Long-term Contextual Effects in Education: Schools and Neighborhoods

Jean-William Laliberté*

Abstract

This paper decomposes total childhood exposure effects – the causal effect of growing up in a better area – into separate school and non-school neighborhood components. To do so, it brings together two research designs. First, I implement a spatial regression-discontinuity design to estimate school effects. Second, I study students who move across neighborhoods in Montreal during childhood to estimate total exposure effects. I find that total exposure effects on educational attainment are large, but that between 50% and 70% of the long-term benefits of moving to a better area are due to access to better schools rather than to neighborhoods themselves.

I Introduction

Improving graduation rates and college attendance are high-priority objectives shared by community leaders, researchers and policymakers. Educational outcomes, however, vary greatly across regions, neighborhoods and schools. Given the sizable economic and nonpecuniary benefits to education, disparities in educational attainment can translate into persistent socio-economic inequality in adulthood.[1] Multiple policy interventions target neighborhoods directly or incentivize families to relocate to low-poverty areas, motivated by the belief that social context significantly influences students' aspirations and learning. Schools are key institutions of local communities, plausibly constituting a pivotal mechanism fueling spatial inequalities. Yet, empirical evidence on the relative importance of schools and of neighborhoods for educational attainment remains scarce, despite the important implications of such information for the allocation of public resources towards policies directed at either schools or neighborhoods. For instance, in many jurisdictions, school attendance is strictly residence-based, making these two dimensions observationally indistinguishable. Disentangling neighborhood and school effects is further complicated by sorting of families; identifying separate causal effects for schools and neighborhoods requires two sources of exogenous variation.

*I am very grateful to Natalie Bau, Kory Kroft and Philip Oreopoulos for their guidance throughout this project. Thanks to Josh Angrist, Gustavo Bobonis, Raj Chetty, David Deming, John Friedman, Nicolas Gendron-Carrier, Michael Gilraine, Hugh Macartney, Richard Mansfield, Ismael Mourifié, Matthew Notowidigdo, Mathieu Marcoux, Juan Morales, Scott Orr, Rob Oxoby, Uros Petronijevic, Marc-Antoine Schmidt, Michel Serafinelli, Aloysius Siow and Alex Whalley for insightful discussions and comments. I also thank Simon Bézy, Latifa Elfassihi and Sophie LeBoutillier from the Ministère de l'Éducation et de l'Enseignement supérieur for their assistance with the data. Financial support from the Social Sciences and Humanities Research Council (SSHRC) and the Ontario Graduate Scholarship (OGS) is gratefully acknowledged. All mistakes are my own. Email: jeanwilliam.lalibert@ucalgary.ca

[1] On the economic returns to education see Oreopoulos and Petronijevic (2013), Moretti (2004), Card (2001), and Angrist and Krueger (1991), and, specifically for Canada, Boudarbat, Lemieux and Riddell (2010). These returns are particularly large for marginal students (Zimmerman, 2014). On non-pecuniary benefits, see Heckman, Humphries and Veramendi (2017), Oreopoulos and Salvanes (2011) and Milligan, Moretti and Oreopoulos (2004).

This paper examines the separate effects of schools and neighborhoods on long-term educational outcomes. In particular, I evaluate the long-term impact of growing up in a better area and calculate the fraction of the benefits that are driven by school quality. To do so, I combine unique student-level administrative data with several key institutional features of Quebec's education system to overcome the stringent data and institutional requirements that have hindered analyses of the separate contribution of schools and neighborhoods. The large longitudinal database follows students who grew up in the region of Montreal throughout their entire educational career and tracks them on a yearly basis as they switch schools, move across neighborhoods, and make higher education investments.

My empirical framework brings together two research designs to develop a new approach that allows me to decompose *total exposure effects* – the combined effect of an additional year of exposure to a given neighborhood and to its schools – into school and non-school neighborhood components. The analyses incorporate variation from one instrument shifting school quality alone (holding neighborhood quality constant) with another shifting both schools and neighborhoods simultaneously. In practice, the decomposition exercise proceeds in several steps. First, I construct measures of school *predicted gains* (school quality) representing school-level differences in educational attainment that cannot be accounted for by where students reside. Second, I estimate the causal effect of school quality on long-term educational outcomes by implementing a regression discontinuity design to recover forecast-unbiased school effects. With these in hand, one can predict how differences in average school quality between two locations causes differences in educational attainment. As a third step, the long-term benefits of moving to a better area earlier (total exposure effects) are estimated using a movers design. Finally, I evaluate what fraction of these total benefits remains after taking out differences in school effects. The results indicate that total exposure effects are large but mostly driven by schools rather than by other non-school neighborhood characteristics.

Quebec is particularly well suited for investigating the role of schools separately from neighborhoods. The province operates two parallel public school systems – one French and one English – thereby allocating neighbors to different default neighborhood schools on the basis of their mother tongue. Importantly, parents are allowed to opt out of these default options, breaking the deterministic link between schools and neighborhoods within language groups. In addition, private schools are widespread and relatively affordable in Montreal, generating further independent variation between schools and neighborhoods. In this paper, I use local variation in the quality of French school default options to instrument for school quality in a spatial regression-discontinuity framework (RD-IV), leveraging the fineness of the spatial information in the administrative files. Crucially, the fact that catchment areas for English and French default schools are different allows me to check the validity of the RD-IV approach by performing placebo tests. I show that the educational outcomes of students attending English schools are smoothly distributed around *French* primary school boundaries, thereby confirming that these boundaries do not coincide with discontinuous changes in non-school unobserved attributes.

Next, I estimate the magnitude of total exposure effects by focusing on movers. To address the endogeneity of location decisions, I exploit variation in the *timing* of within-city moves across families.[2] Intuitively, if exposure matters, the educational outcomes of movers should converge towards those of the permanent residents of their destination (children who always resided in the same area) with increasing time spent in that location. The reduced-form object of interest is a convergence rate. To insure the estimates are not

[2] Aaronson (1998) and Weinhardt (2014) also use variation from movers to identify neighborhood effects. Chetty and Hendren (2018*a*,*b*) use rich tax data to track families with children who move across commuting zones and counties in the U.S. and estimate the causal effect of places on earnings. Similar identification strategies have also been used to analyze health care utilization (Finkelstein, Gentzkow and Williams, 2016), physician practice style (Molitor, 2018), the impact of EITC on labor supply (Chetty, Friedman and Saez, 2013), and brand preferences (Bronnenberg, Dubé and Gentzkow, 2012).

confounded by sorting into different areas, the model relies on comparisons between children who started in the same neighborhood and moved to the same neighborhood, but did so at different ages. The main empirical specification includes both origin-by-destination and age-at-move fixed effects, with the identifying assumption that the degree of selection into locations does not vary systematically with children's age at the time of the move. To provide support for this assumption, I conduct a series of robustness checks, notably family fixed effects specifications and controlling for time-varying observables around the time of the move.

My estimates suggest that movers' educational attainment converges linearly at an annual rate of about 4.5% towards the outcomes of the permanent residents in their destination neighborhood. Put differently, moving one year earlier to a place where permanent residents have 1 more year of education than those of one's origin location increases own educational attainment by 0.045 years. Extrapolating over 10 years of compulsory schooling, these effects account for almost half (45%) of the differences in outcomes of permanent residents between origin and destination. The magnitude of these effects is remarkably similar to that reported in Chetty and Hendren (2018*a*) despite important differences between the two settings, notably my emphasis on smaller geographic units and within-city variation.

To decompose these total exposure effects into school and non-school components, I partition the mean outcomes of permanent residents of a given neighborhood into a part reflecting the average quality of local schools and a non-school residual.[3] With forecast-unbiased measures of school quality in hand, one can predict by how much movers' outcomes would improve on the basis of spatial differences in school quality alone, and evaluate the share of total gains of moving that can be accounted for by these school effects. I find that between 50% and 70% of the benefits of moving to a better area are due to access to better schools. In other words, once we take out differences in causal school effects between origin and destination, the rate of convergence on the remaining differences associated with non-school factors is less than half the size of the total convergence rate of 4.5%. This suggests that if average school quality was equalized across all neighborhoods, the gains associated with growing up in a relatively better area would be more than halved. Even in a context where students have the option to opt out of their local educational institutions, causal place effects are driven for the most part by schools rather than by neighborhoods. Nonetheless, a residual neighborhood exposure effect persists above and beyond the contribution of schools.

This paper brings together several literatures. First, it speaks directly to the classic question "Do neighborhoods matter?". Correlational analyses generally find strong associations between neighborhood poverty and success at school (Sharkey and Faber, 2014; Burdick-Will et al., 2011). In contrast, most experimental and quasi-experimental studies that tackle the challenging task of isolating place effects from non-random sorting of families into neighborhoods have found limited evidence of static neighborhood effects on educational and economic outcomes (Ludwig et al., 2013; Kling, Liebman and Katz, 2007; Oreopoulos, 2008, 2003; Jacob, 2004).[4] In a recent re-analysis of the Moving to Opportunity (MTO) experiment, Chetty, Hendren and Katz (2016) show that children do benefit from moving to better locations both in terms of earnings and college enrollment, but that these gains only materialize for youth who moved before the age of 13, consistent with cumulative exposure effects. Similarly, Chetty and Hendren (2018*a*,*b*) estimate large exposure effects for children moving across U.S. commuting zones. Given that school attendance is

[3] This empirical approach is consistent with a simple education production function in which cumulative school and non-school neighborhood inputs are additively separable and treatment effects are constant across students. Analyses presented in Sections II and IV suggest these functional form assumptions are plausible in my setting. Section III describes the corresponding conceptual framework in details.

[4] Notable exceptions include Goux and Maurin (2007) who find positive effects of neighborhood peers on the probability of repeating a grade using variation from public housing projects in France, and Gould, Lavy and Paserman (2011) who find significant effects of childhood conditions on adult outcomes in Israel.

generally residence-based, these estimates of neighborhood exposure effects also reflect differences in local school quality (Altonji and Mansfield, 2014). My estimates of total exposure effects are consistent with this prior literature. However, my main focus is on unpacking the role of schools as a mechanism: accounting for the cumulative nature of contextual effects, I demonstrate that neighborhood exposure effects operate mostly via schools.

Second, my paper relates to a parallel stream of research evaluating the causal impact of schools on educational and labor market outcomes. Large effects of attending a better school are found using quasi-experiments (Gould, Lavy and Paserman, 2004), lottery-based designs (Angrist et al., 2017; Deming et al., 2014; Dobbie and Fryer, 2015, 2011), and admission threshold rules (Pop-Eleches and Urquiola, 2013; Jackson, 2010). Using similar research designs, however, Abdulkadiroğlu, Angrist and Pathak (2014) and Cullen, Jacob and Levitt (2006) respectively find no positive effects of attending an elite school or of attending one's preferred school in a school choice program. My paper takes a different approach, exploiting spatial discontinuities in the spirit of Black (1999), to show the early schooling environment has a long-term impact: residing on the better side of a French primary school boundary at age 6 affects educational outcomes measured more than 10 years later.

I also contribute to a growing body of research contrasting the magnitude of school and neighborhood effects.[5] Historically, researchers have generally focused on either schools or communities, with a few review papers speculating on the relative effectiveness of school and neighborhood interventions by comparing separate studies. Fryer and Katz (2013) and Katz (2015) contrast results from the MTO experiment (Ludwig et al., 2013; Kling, Liebman and Katz, 2007), which induced low-income families to move to low-poverty neighborhoods, with the effects of the Harlem Children's Zone experiment, which combines both school-level and community-level interventions (Dobbie and Fryer, 2015, 2011). They conclude that school interventions are likely more effective than community programs for educational outcomes, a conclusion also reached by Oreopoulos (2012). At larger levels of aggregation, Rothstein (2017) finds differences in quality of K-12 education (measured by test scores) account for little of between-city differences in intergenerational mobility, while Card, Domnisoru and Taylor (2018) find that state- and county-level school quality was a key factor driving regional differences in upward mobility in the early 20th century. My paper adds to this evidence by directly separating school from neighborhood effects using two instruments simultaneously, showing that school quality goes a long way explaining *why* place matters for educational attainment.

The methods used in this paper have several empirical benefits. Movers likely constitute a more diverse cross-section of the population than samples of experimental studies that focus on very disadvantaged households (Oreopoulos, 2003) or negatively-selected populations of lottery applicants (Chyn, 2016), contributing to the external validity of the results. Also, by using outcome-based measures of neighborhood and school quality, I circumvent the issue of choosing which observable characteristics to use to proxy for quality. For instance, school input measures and teacher observable characteristics often fail to predict effectiveness, despite the evidence that both schools and teachers have large causal effects on student outcomes (Dobbie and Fryer, 2013; Chetty, Friedman and Rockoff, 2014*b*; Rivkin, Hanushek and Kain, 2005; Hanushek, 1986).

The rest of the paper proceeds as follows. First, I describe the institutional context and the data in Section II. Then, to fix ideas and motivate the empirical analyses, I set up a conceptual framework in

[5]In sociology, Carlson and Cowen (2015) examine short-run variation in test score gains across schools and neighborhoods in Milwaukee, and Wodtke and Parbst (2017) explore how school poverty mediates neighborhood effects on math and reading tests in the PSID. Sykes and Musterd (2011) study how school characteristics mediate the relationship between neighborhood characteristics and test scores in the Netherlands. In economics, Card and Rothstein (2007) separately examine the effects of school and neighborhood segregation on test scores, and Billings, Deming and Ross (2016) consider the role of school and neighborhood peers in the formation of criminal networks.

Section III. Econometric models for each step of the empirical procedure and the associated results are shown in Section IV. A host of robustness checks are conducted in Section V, and Section VI concludes.

II Data and Background

II.A Quebec's Education System

Levels of education. In Quebec, education is compulsory from age 6 to 16, and most children enroll in kindergarten at age 5. Children complete six years of primary education (grades 1-6), and then attend a secondary school for five more years (grades 7-11), until obtaining a secondary school diploma (*diplôme d'études secondaires*, *DES*), or equivalent qualifications. Grade repetition is common and over 20% of students drop out of secondary school before obtaining any degree.

The higher education system differs considerably from standard North American systems.[6] In Quebec, there is a sharp hierarchical distinction between *college* and *university*, the former being a pre-requisite for the latter. After secondary school, most students enroll in college in either a pre-university (2 years) or technical program (3 years). The typical student obtaining a pre-university college degree then enrolls in a 3-year bachelor degree program in university. In the empirical application, I measure neighborhood and school exposure up until the academic year a student is aged 15 on September 30, inclusive. All educational investments made after that point are considered outcomes.

School choice between sectors. Quebec's education system possesses multiple elements of school choice that contribute to breaking the mechanical link between area of residence and school attendance. At the primary and secondary levels, two public school systems operate in parallel – one French and one English. Public schools are governed by schools boards, which are responsible for personnel, transportation, infrastructure, and the allocation of resources across schools. School boards are language-specific, with every residential address falling within the territory of one English and one French school board.[7] Importantly, the attendance zones of English and French schools are not the same. Hence, two neighbors with different mother tongues who both attend their nearest language-specific public school likely have school peers who originate from different neighborhoods. Access to instruction in English is restricted to anglophones born in Canada, a strictly enforced rule where parents must obtain an eligibility certificate before enrolling a child in an English school. In the language of the law, anglophones are students whose mother or father attended an English primary or secondary school in Canada. Almost all immigrants are *de facto* forbidden from attending English schools. Exceptions to the rule are rare.[8]

In comparison with other Canadian provinces and the U.S., private schools are widespread in Quebec, notably at the secondary school level. In Montreal, almost a third of all students attend a private secondary school. Private schools do not have attendance zones and are relatively accessible given generous subsidies. Subsidized private schools are also subject to the language of instruction restriction.[9]

[6] Figure A1 shows the typical education course towards a bachelor degree in Quebec and in a standard North American system. No transition between levels of education in Quebec coincide with the age at which students transition in other educational systems. The number of years of education associated with a bachelor degree, however, remains the same.

[7] Before 1998, school boards were religion-specific (Catholic or Protestant), but individual schools were still either French or English.

[8] Language restrictions do not apply to post-secondary institutions.

[9] Non-subsidized English schools are allowed to enroll non-English speaking students. However, these schools are uncommon and represent less than 1% of total enrollment (Duhaime-Ross, 2015). Among subsidized private schools, very few charge the maximum fee allowed by law (Lefebvre, Merrigan and Verstraete, 2011). A minority of these schools have entrance exams, yet the vast majority of students taking such exams are admitted to their preferred school (Lapierre, Lefebvre and Merrigan, 2016).

School choice within sectors. Quebec's open enrollment policy stipulates that parents have the right to enroll their child in the school of their choice (*libre choix*), subject to capacity constraints and language restrictions. In practice, school boards assign children to default neighborhood schools, and parents who desire to enroll their child in a public school other than the one they are assigned must complete the relevant paperwork at the neighborhood school. Default options may induce two sets of parents living in the same area to enroll their children in different schools since catchment area boundaries often cut through neighborhoods. In this paper, I focus exclusively on French primary school boundaries since English primary school boundaries are not as well defined – some English schools offer different programs (e.g. English Core vs. Bilingual) and their catchment areas often vary by program. French primary school boundaries serve as the basis for a regression-discontinuity analysis described in Section IV.B.[10]

Importantly, over the time period studied here, there existed no public information about relative primary school quality and performance such as rankings on outcome-based measures. All Montreal school boards strongly oppose public disclosure of rankings or quality indicators at the primary school level.[11] If enrollment exceeds capacity, priority is given to children residing in the school's catchment area and to siblings of children attending the school, and students opting-out of their assigned school are not eligible for school bus transportation. Other non-residence based admission criteria are used for elite magnet schools. The neighborhood school therefore acts as a default option, and catchment area boundaries as cost shifters. In my data, every neighborhood school enrolls at least some students residing outside its catchment area.[12]

II.B Data

The main data consist of student-level administrative records provided by Quebec's Ministry of Education covering all levels of education (primary school to university). Separate files from four different branches of the Ministry were matched using unique student identifiers. For each year students are enrolled in primary and secondary education, school attended, grade level, and the six-digit postal code of residence are recorded. Postal codes (very small geographic areas, generally equivalent to a block-face or a unique apartment building) determine the default neighborhood schools (one English and one French). Catchment areas were manually geocoded on that basis.

In addition to the assigned neighborhood schools, I calculated for each postal code the distance to the nearest catchment area boundary, distance to the nearest public school, associated census tract and Forward Sortation Area (FSA; postal-code-based neighborhoods constituting the main geographic unit of analysis).[13] All distances were calculated separately for the English and French public school systems. In Montreal, students reside in over 500 different census tracts and about 100 FSAs. For confidentiality reasons, school identifiers and six-digit postal codes were de-identified after the relevant distance variables were calculated.

[10] Appendix Figure A2 shows that French primary school boundaries often cut through census tracts. At the secondary school level, English public schools in Montreal do not have catchment areas, but French public schools do.

[11] Secondary school rankings are published yearly in mainstream media. In 2015, a well-known newspaper published partial rankings of Montreal public primary schools for the first time. The cohorts of students analyzed in this paper had left primary school many years before that.

[12] One important reason why capacity constraints do not appear to be binding is that Quebec's school system was experiencing a decline in school-age population over the time period covered here, which notably led to several public school closures in the early 2000s.

[13] A FSA is defined by the first three digits of a postal code. FSAs are regularly used to operationalize neighborhoods in Canadian research (Card, Dooley and Payne, 2010) and are sometimes used for determining eligibility to community programs. For example, the *Pathways to Education* program targeting residents of the neighborhood of Pointe-Saint-Charles is available to households living in one specific FSA. In my data, educational attainment is relatively smoothly distributed spatially within FSAs: only 3% of the within-FSA residual variation occurs between census tracts. Online Appendix E examines the robustness of exposure effects estimates to using census tracts instead of FSAs to define neighborhoods.

Student demographics – age on September 30, gender, mother tongue, country of origin, language spoken at home – are included, in addition to time-varying variables such as school day care use (primary school only) and an indicator of whether a student is currently considered to have learning difficulties (primary and secondary school).[14] In addition, I append census tract-level characteristics from the 2001 Canadian Census.

In terms of long-term educational outcomes, the data include enrollment and graduation information for secondary school, and all vocational, college, and university programs. I use these to calculate – among other outcomes – university enrollment, timely secondary school graduation, and number of years of education. More detailed information regarding the construction of the outcome variables is provided in the Data Appendix.

The sample is focused on residents of the Island of Montreal, Quebec's most populous region and main urban center. This territory encompasses three francophone and two anglophone school boards, and includes the city of Montreal and a few smaller municipalities located in the suburban westernmost part of the Island or enclaved within the city of Montreal. These municipal divisions are irrelevant for school resources administration purposes.

Administrative records were obtained for five cohorts of students who started primary school between 1995 and 2001, following students until the 2014-2015 academic year.[15] The sample consists of all students who resided on the Island of Montreal when entering grade 1 (100,929 students). This selection rule ensures that every student's entire education history is known, and therefore excludes students who moved to Montreal after completing grade 1 elsewhere. On average, about 210 students per FSA enter grade 1 on any given year. The main sample (92,764 students) excludes all students who left Quebec's education system before turning 16.[16]

II.C Descriptive and Summary Statistics

Descriptive statistics by mobility status are shown in Table A1. Permanent residents are defined as those who, by the age of 15, had always resided in the same FSA. I distinguish between movers who were still living in Montreal by age 15 , and those who had moved off the Island but remained in the province. Because of the within-city focus of this paper, students who left Montreal are excluded from the empirical analyses.

In Montreal, students are on average 6 years old when they enter primary school. Only half the sample consists of native French speakers, but 75% of students attend school in French. *Allophones* – defined as individuals whose mother tongue is neither French nor English – make up almost a third of the sample. Nevertheless, the vast majority of students are born in Canada (90%). Anglophones are overrepresented among permanent residents, while francophones are disproportionately more likely to move outside of Montreal, and allophones to move within Montreal. At baseline (in grade 1), 4% of students are considered to have learning difficulties (flagged "in difficulty"), and the fraction increases sharply over time. By the time they reach the age of 15, almost a third of the sample will have been flagged at least once. In general, movers appear to be negatively selected: In grade 1, 3% of permanent residents are in difficulty, while 5% of movers

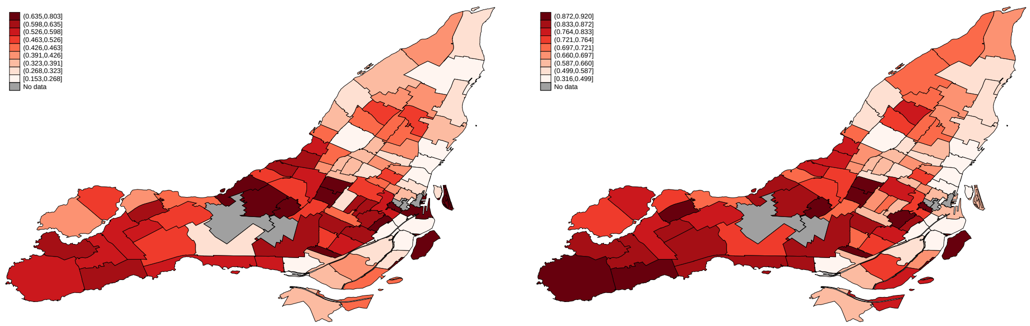
[14]On any given year in primary and secondary school, students with social maladjustment or learning disabilities can be identified as being "in difficulty". School boards receive extra funding to support these students, and many observers worry that schools may 'over-diagnose' students as a result. Yet, the predictive power of this variable with respect to educational attainment is stunning. The probability that one obtains a secondary school diploma on time decreases monotonically with each year flagged in difficulty (Figure A3). For the two earliest cohorts, the probability of obtaining a bachelor degree is 36% for students never identified in difficulty, while it is only 5% among those who were flagged at least once.

[15]Data for the 1997 and 1999 cohorts are not available.

[16]In primary and secondary school, attrition is generally due to students leaving the province. In exceptional cases, some students may disappear from administrative records if they attend illegal schools, or are home schooled. Students leaving the system are disproportionately non-French speakers and immigrants. See the Data Appendix for more details on attrition.

Figure 1: Spatial Variation in Educational Outcomes

University enrollment rates | Timely graduation in secondary school



Notes: Statistics based on permanent residents. Outcomes are adjusted for cohort effects. Data for FSAs with fewer than 10 permanent residents are not shown (no data).

are.

The number of years for which I track students varies across cohorts, hence observed educational attainment is higher for earlier cohorts, by construction. Appendix Table A2 reports summary statistics for some educational outcomes separately by cohort. Roughly 76% of students obtain a secondary school diploma (*DES*), but only 61% do so on time (in five years), with little variation across cohorts. The college enrollment rate is consistent across cohorts, at 70%. In terms of university-level outcomes, as of 2015, 46% of students who started primary school in 1995 had enrolled in university and 28% had completed a bachelor degree. Virtually no student of the 2001 cohort has a bachelor degree yet, but 22% of them are enrolled in university. Every econometric model in this paper includes cohort fixed effects to account for these differences.

Educational attainment varies dramatically across neighborhoods of Montreal. Figure 1 maps differences in mean educational attainment of permanent residents across FSAs.[17] The gap between neighborhoods with best and worst outcomes is abysmal, with fractions of students completing high school on time ranging from 32% to 92%. The gap grows even larger in terms of university enrollment, with a minimum rate of 15% and a maximum of 80%. Even starker disparities emerge across census tracts (Figure A7).

School attendance. Given the variety of school choice options available in Montreal, students living in the same neighborhood need not attend the same school. For instance, at baseline, students living in the average FSA attend as many as 57 different primary schools.[18] When entering grade 1, 63% and 50% of students in French and English schools attend their neighborhood school, respectively. In total, 41% of students opt-out of their default option at baseline, with this proportion exceeding 70% by the end of secondary school.[19] Opt-out rates vary between the primary and secondary school levels primarily because of differences in availability of private school options. Around 12% of Montreal students are in the private sector in primary school, a fraction rising to almost 30% in secondary school (Figure A6). Yet, geography remains an important factor for many parents when it comes to deciding which school their child will attend. For example, among students in French schools at baseline, 68% attend their default school if that school is the nearest French public school from their house, while only 50% do so if it is not.[20] Despite the open enrollment policy, there

[17] Values of $\bar{y}_n^{PR}$ are adjusted for cohort differences. Mean outcomes are neighborhood fixed effects from a regression of outcome y on neighborhood and cohort fixed effects. The fixed effects are then re-centered so that their average is equal to the unconditional mean $\bar{y}$.

[18] See Appendix Figure A5 for distribution of FSAs by number of schools.

[19] By definition, students in English secondary schools all opt-out since English public schools do not have attendance zones at the secondary level. At the age of 15, 58% of students in French schools attend a school other than their default option.

[20] For 30% of students in French schools the default option is not the nearest French public primary school.

is a large discontinuity in school attendance around boundaries at baseline, suggesting that many parents passively select the default option. This is consistent with a body of evidence in behavioral economics and psychology on the importance of default options (Chetty, 2015; Lavecchia, Liu and Oreopoulos, 2014).[21]

III Conceptual Framework

This section lays out a human capital production function that linearly incorporates cumulative school and neighborhood non-school inputs, expanding the framework of Chetty and Hendren (2018*b*) to allow for multiple contextual dimensions. It first describes the outcomes of permanent residents and movers parsimoniously in terms of school and neighborhood quality. These expressions map to the parameters I identify with the two empirical strategies. Finally, how a decomposition of total exposure effects can be achieved is discussed.

Education production function. Consider a framework in which educational investment in children takes place over compulsory schooling years (up to year A) and a long-term outcome is realized and measured after the investment years. The education production function is cumulative and separately additive in family, school, and neighborhood (non-school) inputs:

$$y_i = \sum_n a_{in} \left[\lambda\mu_n + \omega\tilde{\psi}_{s(n(i))}\right] + A\tilde{\theta}_i \tag{1}$$

where y_i is a measure of educational attainment for child i, a_{in} is the number of years the child resided in location n (with $\sum_n a_{in} = A$), $\tilde{\psi}_{s(n(i))}$ denotes the annual average quality of schools attended while residing in location n, and $\tilde{\theta}_i$ are annual average family inputs. Neighborhood and school quality are denoted by variables μ_n and $\tilde{\psi}_{s(n(i))}$, and parameters λ and ω respectively represent the causal effect of one year of exposure to better non-school neighborhood amenities and the causal effect of attending a better school for one year.[22] Note that average school quality $\tilde{\psi}_{s(n(i))}$ remains indexed by i because students living in the same area can attend different schools.

School effects. To isolate school effects from any neighborhood-related variation, I focus on the subsample of **permanent residents (PR)** – children who always resided in the same place – for whom $a_{ik} = A$ for neighborhood n, and $a_{ik} = 0$ for all other locations $k \neq n$. In practice, school quality $\tilde{\psi}_{s(n(i))}$ and neighborhood quality μ_n are unobserved. Hence, I partition PRs' educational attainment into measurable school-related and neighborhood-related terms, and an idiosyncratic residual ν_i unrelated to either schools

[21] Discontinuity in school attendance at French primary schools boundaries is shown in Figure A4. To create the figure, I randomly pick one of the two schools that share a boundary, separately for each boundary. I then calculate the fraction of students who attend the randomly chosen school as a function of distance to the nearest boundary. Students at positive distances are residing in the catchment area of the randomly chosen school. On the left side of the border (negative distances), 20% of students attend the school located on the other side, rather than their own default option or any other French school.

[22] To keep the model tractable, I do not explicitly include disruption costs associated with moving or switching school in the production function. The empirical model developed in Section IV accounts for any age-variant disruption costs with the inclusion of age-at-move fixed effects. Also, the framework abstracts from school and neighborhood effects that are not proportional with length of exposure. For instance, any benefits associated with residing close to a university at the age of 20, which are unrelated to how old one was when one arrived at that location, are not taken into account. This is mainly because in practice any such effects are absorbed by origin-by-destination fixed effects when estimating total exposure effects, and the inclusion of these fixed effects is necessary for identification. However, such exposure-independent effects may not be that large in a city like Montreal, where numerous universities and colleges are scattered across the city. In addition, Chetty and Hendren (2018*a*) find that place effects converge toward zero as age-at-move approaches 23, and Chetty, Hendren and Katz (2016) find no benefits of moving to a low-poverty area in adulthood in the MTO experiment.

or neighborhoods:

$$y_i^{PR} = \Omega_{s(n(i))} + \Lambda_n + \nu_i \tag{2}$$

where $\Omega_{s(n(i))}$ reflects both cumulative causal school effects over student i's childhood $A\omega\tilde{\psi}_{s(n(i))}$ as well as average sorting into schools, and Λ_n is defined accordingly for neighborhood non-school amenities. Put differently, $\Omega_{s(n(i))}$ is a biased measure of true school effects because it incorporates the partial correlation between school quality $\tilde{\psi}_{s(n(i))}$ and parental inputs $\tilde{\theta}_i$.

My first empirical objective is to obtain a consistent estimate of π – a coefficient that measures the relationship between estimated school effects $\Omega_{s(n(i))}$ and causal school effects $A\omega\tilde{\psi}_{s(n(i))}$ – in order to recover forecast-unbiased measures of school predicted gains $\pi\Omega_{s(n(i))}$ (Chetty, Friedman and Rockoff, 2014*a*).[23] This can be achieved by using a valid instrumental variable that exogenously shifts school quality independently of neighborhood quality ("first-stage") and that is uncorrelated with parental inputs ("exclusion restriction"). An OLS regression of y_i^{PR} on $\Omega_{s(n(i))}$ and a set of neighborhood fixed effects yields a coefficient on $\Omega_{s(n(i))}$ of one, by construction. In contrast, instrumenting for $\Omega_{s(n(i))}$ yields a regression coefficient equal to π. Letting $\bar{x}_n^{PR}$ denote the average of some variable x for PRs of neighborhood n, $\pi\bar{\Omega}_n^{PR}$ then provides a forecast-unbiased measure of the average cumulative causal school effects for PRs of neighborhood n, $A\omega\bar{\psi}_n^{PR}$.

Total exposure effects. The total effect of growing up in a given area incorporates both school and non-school neighborhood inputs. To obtain causal estimates of these *total exposure effects*, I rely on **movers.** For one-time movers, let $o(i)$ denote the origin neighborhood of mover i, $d(i)$ denote the destination, and m_i the age at which student i moved. For these students, $a_{io} = m_i - 1$, $a_{id} = A - (m_i - 1)$, and $a_{ik} = 0$ $\forall k \neq o, d$. Their educational attainment is given by

$$y_i = A\left[\lambda\mu_{d(i)} + \omega\tilde{\psi}_{s(d(i))} + \tilde{\theta}_i\right] - (m_i - 1)\underbrace{\left[\lambda\left(\mu_{d(i)} - \mu_{o(i)}\right) + \omega\left(\tilde{\psi}_{s(d(i))} - \tilde{\psi}_{s(o(i))}\right)\right]}_{\text{Total exposure effects } (e_i(o,d))}. \tag{3}$$

Equation (3) highlights that the long-term outcomes of movers depend on the quality of schools and neighborhoods in both places as well as on the length of *exposure* to each place, which varies with age-at-move. Total exposure effects $e_i(o,d)$ are the gains of living in and attending schools of area d for one year relative to area o. To take equation (3) to the data, it is useful to re-write it as a function of variables that can be readily measured. For instance, since permanent residents' mean outcomes are a function of the same contextual inputs, the total effect of living one year in area d relative to area o is directly related to the outcomes of PRs in both locations. Substituting the difference in average outcomes between permanent residents of neighborhoods d and o, $\Delta\bar{y}_{od} = \bar{y}_d^{PR} - \bar{y}_o^{PR}$, into equation (3):

$$e_i(o,d) = \left(\frac{1}{A}\right)\Delta\bar{y}_{od} + \left(c_{i(o,d)} - 1\right)\omega\left[\bar{\psi}_d^{PR} - \bar{\psi}_o^{PR}\right] - \left[\bar{\theta}_d^{PR} - \bar{\theta}_o^{PR}\right] \tag{4}$$

where $c_{i(o,d)} = \frac{\tilde{\psi}_{s(d(i))} - \tilde{\psi}_{s(o(i))}}{\bar{\psi}_d^{PR} - \bar{\psi}_o^{PR}}$ is a school compliance factor indicating the propensity of movers to attend schools of comparable quality to those attended by permanent residents in their origin and destination. Positive exposure effects imply that the cumulative gains of moving to a $\Delta\bar{y}_{od}$-unit better area should grow

[23] Put differently, π denotes the fraction of the effect of $\Omega_{s(n(i))}$ on y_i^{PR} reflecting causal variation. More details about the interpretation of π are provided in Online Appendix C.

(shrink) with the amount of time spent the destination (with age-at-move). Empirically, the magnitude of annual exposure effects can be assessed by comparing students who moved from the same origin to the same destination at different ages. Intuitively, if d is "better" than o, then a student who moved at age 9 is expected to have better outcomes than her peer who made the same move at age 12 since she will have been exposed to the better area for three more years.[24]

Empirically, the object of interest is the rate at which movers' outcomes *converge* towards those of the permanent residents of their destination with the number of years of exposure to that location, which can be estimated by regressing movers' outcomes y_i on the interaction term $(m_i - 1) \times \Delta\bar{y}_{od}$. This convergence rate, denoted by β, corresponds to $\frac{Cov(e_i, \Delta\bar{y}_{od})}{Var(\Delta\bar{y}_{od})}$. Equation (4) indicates that its magnitude is decreasing in the degree of sorting of permanent residents (e.g. the extent to which variation in $\Delta\bar{y}_{od}$ reflects differences in $\bar{\theta}_d^{PR} - \bar{\theta}_o^{PR}$).[25] Convergence increases with the compliance factor $c_{i(o,d)}$. Under full school compliance ($c_{i(o,d)} = 1 \forall i$) and no sorting of permanent residents, the convergence rate is equal to $\frac{1}{A}$. A non-zero convergence rate indicates that there are benefits to moving to a better area, but does not necessarily imply that neighborhoods matter independently of schools. If neither schools nor neighborhoods matter (i.e. $\lambda = \omega = 0$), then the convergence rate is zero.

Decomposition. Total exposure effects encompass both changes in school and non-school neighborhood inputs. My main goal is to decompose these gains into a school component reflecting differences in causal schools effects across locations and a non-school component attributable to other neighborhood amenities. Among permanent residents, spatial differences in school quality cause educational attainment to be higher by $\pi\Delta\Omega_{od} = \pi\left(\bar{\Omega}_d^{PR} - \bar{\Omega}_o^{PR}\right)$ in location d relative to o. Taking out these differences, the remaining gap, $\Delta\bar{y}_{od}^{-s} \equiv \Delta\bar{y}_{od} - \pi\Delta\Omega_{od}$, reflects non-school factors. With measures of $\Delta\Omega_{od}$ and π in hand, one can decompose total exposure effects accordingly:

$$e_i(o,d) = \left(\frac{1}{A}\right)\Delta\bar{y}_{od}^{-s} + \left(\frac{c_{i(o,d)}}{A}\right)\pi\Delta\Omega_{od} - \left[\bar{\theta}_d^{PR} - \bar{\theta}_o^{PR}\right]. \tag{5}$$

From an accounting perspective, the total convergence rate can similarly be separated into school and non-school components:

$$\beta = \underbrace{\frac{Cov(e_i, \Delta\bar{y}_{od}^{-s})}{Var(\Delta\bar{y}_{od}^{-s})}\frac{Var(\Delta\bar{y}_{od}^{-s})}{Var(\Delta\bar{y}_{od})}}_{\beta^{non-school}} + \underbrace{\frac{Cov(e_i, \pi\Delta\Omega_{od})}{Var(\pi\Delta\Omega_{od})}\frac{Var(\pi\Delta\Omega_{od})}{Var(\Delta\bar{y}_{od})}}_{\beta^{school}}. \tag{6}$$

I define the share of total effects accounted for by causal school effects to be the ratio of the school component over the total convergence rate

$$S^{school} \equiv \frac{\beta^{school}}{\beta} = \frac{Cov(e_i, \pi\Delta\Omega_{od})}{Cov(e_i, \Delta\bar{y}_{od})}. \tag{7}$$

[24] If individual inputs adjust in response to changes in other inputs, then the effect of moving a student across neighborhoods should be interpreted as a policy effect which encompasses parental responses (Todd and Wolpin, 2003). For instance, prior research suggests that parental effort and school quality are treated as substitutes (Pop-Eleches and Urquiola, 2013; Houtenville and Conway, 2008).

[25] This is under the assumption that families with high unobservable characteristics select into better schools and neighborhoods: $Cov(\lambda\mu_n + \omega\bar{\psi}_n^{PR}, \bar{\theta}_n^{PR}) > 0$. If the parents of students with low family inputs are more likely to sort into better schools and neighborhoods, then the convergence rate increases with sorting of permanent residents.

The *school share*'s magnitude decreases with the amount of sorting of permanent residents into schools. It increases with the compliance factor $c_{i(o,d)}$ and therefore depends on the school choice regime analyzed. Also, S^{school} increases linearly with π, and is effectively equal to zero if $\pi = 0$. Under full school compliance ($c_{i(o,d)} = 1 \forall i$) and no sorting of permanent residents, the school share is the fraction of the total variance $Var\left(\Delta \bar{y}_{od}\right)$ that is attributable to schools, that is $\frac{Cov(\Delta\Omega_{od}, \Delta\bar{y}_{od})}{Var(\Delta\bar{y}_{od})} = \frac{Var(\Delta\Omega_{od}) + Cov(\Delta\Omega_{od}, \Delta\bar{y}_{od}^{-s})}{Var(\Delta\bar{y}_{od})}$.

The education production function used in this paper has several restrictions. Firstly, neighborhood and school effects are both linear in years of exposure.[26] In Section IV.C, I show that movers' outcomes indeed converge linearly to those of permanent residents of the destination. Secondly, the model imposes additive separability of schools and neighborhoods, thereby ruling out complementarities between the two dimensions. I provide evidence that there is no systematic interaction between school and neighborhood quality in my data in the next section. Also, additive separability of schools and neighborhoods is relatively standard in the literature (Gibbons, Silva and Weinhardt, 2013; Card and Rothstein, 2007), and is consistent with results from the Harlem Children's Zone (Fryer and Katz, 2013). Finally, school and neighborhood effects are assumed to be constant across students, an assumption that is also common to most work on school (Deming, 2014), teacher (Chetty, Friedman and Rockoff, 2014*b*) and college (Hoxby, 2015) value-added.

IV Empirical Roadmap

This section describes the econometric specifications used to obtain each empirical object necessary to implement the decomposition of interest, and report the associated estimates. The decomposition procedure is as follows. First, (possibly biased) measures of school quality are obtained. In a second step, the causal effect of an increase in school quality is estimated to recover forecast-unbiased estimates of school effects. The third step corresponds to estimating total exposure effects using the movers design. In the fourth and final step, the school share of total exposure effects is calculated. This is accomplished by first taking out differences in causal school effects between origin and destination (using the estimates recovered in step 2) and evaluating what fraction of the total gains of moving to a better area remain.

IV.A Schools and neighborhoods: Measurement

To obtain measures of school quality, I estimate a two-way fixed effects model on the subsample of permanent residents. The estimating equation is

$$y_{insc} = \boldsymbol{\delta}_{s(i)} + \delta_n + \delta_c + \epsilon_{insc} \tag{8}$$

where y_{incs} is a long-term educational outcome for student i from cohort c, living in neighborhood n and attending the set of schools $s(i)$. The model includes cohort (δ_c), FSA (δ_n) and school ($\boldsymbol{\delta}_{s(i)}$) fixed effects. Intuitively, this model is identified because the set of students living in the same area attend a variety of different schools, and students in the same school reside in different neighborhoods.[27] Since students generally attend two different schools during childhood – one primary and one secondary school – I parameterize the vector of school effects to include a fixed effect for primary school attended at baseline (δ_s^P) and a fixed

[26] Angrist et al. (2017), Dobbie and Fryer (2013), Abdulkadiroğlu et al. (2011) and Autor et al. (2016) also assume that school effects are proportional with number of years. Chetty and Hendren (2018*a*) make a similar assumption for place effects.

[27] Just like models of worker and firm fixed effects are identified from switchers (Abowd, Kramarz and Margolis, 1999), this model requires that students of a given neighborhood be observed in multiple schools and that students from a given school be observed in multiple neighborhoods.

Table 1: Variation Across Neighborhoods and Schools

	Outcome					
	DES in 5 years		University enrollment		Years of education	
	(1)	(2)	(3)	(4)	(5)	(6)
Student-level standard deviation of fixed effects:						
Schools	0.270	0.264	0.249	0.235	1.207	1.141
Neighborhoods (FSAs)	0.138	0.046	0.139	0.062	0.680	0.258
Dependent variable summary statistics:						
Mean	0.706		0.443		13.228	
Standard deviation	[0.456]		[0.497]		[2.113]	
Fixed effects estimated						
Separately	x		x		x	
Simultaneously		x		x		x
Number of students	44,912					
Number of primary schools	440					
Number of secondary schools	218					
Number of neighborhoods	95					

Notes: Sample restricted to permanent residents. School fixed effects are measured by the sum of a primary and a secondary school fixed effect. In columns (1), (3) and (5), school and neighborhood effects are respectively estimated in separate regressions. In columns (2), (4) and (6), all fixed effects are estimated simultaneously from equation (8).

effect for secondary school attended at age 15 (δ_s^S). I therefore obtain a proxy for school quality for each school in the data set. Note that these measures of school quality are net of neighborhood fixed effects and therefore reflect the contribution of schools (and sorting into schools) that cannot be accounted for by where schools gather their students from. Primary school quality is net of the secondary schools its students will eventually attend, and secondary school quality is net of the primary schools it gathers its students from. These outcome-based measures of school quality can be interpreted as predicted gains and reflect any observed and unobserved differences in productive school inputs – e.g. teacher and principal quality. Traditional measures of school quality based on test scores may not fully capture other important dimensions of school effectiveness for long-term educational attainment, such as effects on non-cognitive skills (Jackson, 2016; Heckman, Stixrud and Urzua, 2006).

To describe the amount of variation in the data, Table 1 reports the student-level standard deviation in school ($\boldsymbol{\delta}_{s(i)}$) and FSA ($\delta_n$) fixed effects for the three main outcomes of interest: university enrollment, finishing secondary school on time (DES in 5 years), and years of education. I first report in columns (1), (3) and (5) the raw variation across school and neighborhood fixed effects, not accounting for variation in the other dimension. These benchmarks reflect the dispersion of neighborhood and school fixed effects estimated in separate regressions. For all three outcomes, the variance across schools is about twice as large as the variance between FSAs.[28] In columns (2), (4) and (6), fixed effects for schools and neighborhoods are estimated simultaneously. While the magnitude of the variation across schools barely changes when FSA fixed effects are included, a large fraction – between 55 and 65 percent – of the raw student-level variation across FSAs is accounted for by school attendance. Nevertheless, this preliminary piece of descriptive evidence suggests that there is independent variation across both schools and neighborhoods that cannot be accounted for by the other dimension. It is worth noting that these estimates may be subject to sampling error, an issue I return to in Section V.C.[29]

[28]This result is not due to the fact that there are fewer FSAs than schools, as the patterns replicate at the census tract level (Table E1). Also, these patterns closely reflect the conclusions of Carlson and Cowen (2015), who focus on the variance in test scores growth across neighborhoods and schools in Milwaukee's open enrollment system.

[29]To maintain the mapping between $\bar{\Omega}_n^{PR}$, $\bar{\Lambda}_n^{PR}$ and $\bar{y}_n^{PR}$ intact (i.e. $\bar{y}_n^{PR} = \bar{\Omega}_n^{PR} + \bar{\Lambda}_n^{PR}$), I work with unadjusted estimates in the main analyses. Appendix Table A3 reports standard deviations of $\boldsymbol{\delta}_{s(i)}$ and δ_n for "shrunk" estimates obtained using empirical Bayes techniques (Kane and Staiger, 2008; Chandra et al., 2016; Best, Hjort and Szakonyi, 2017). Such an adjustment

I use the fixed effect estimates reported in column (6) to document two additional stylized facts. Firstly, the student-level correlation between school and FSA fixed effects for years of education is small but positive (0.17), indicating that students residing in better neighborhoods attend better schools on average.[30] Secondly, I follow the approach developed in Card, Heining and Kline (2013) to examine whether there are systematic interactions between the two contexts. Figure A8 is constructed by slicing the distributions of school and FSA fixed effects into deciles, and then plotting the average residuals in each school-by-neighborhood decile cell. Most average residuals are smaller than 0.1 year of schooling, or less than 5% of a standard deviation in the sample of permanent residents. If there were positive interactions between school and neighborhood quality, one would expect abnormally large and positive mean residuals for cells corresponding with high or low deciles in both dimensions. The figure shows no such discernible pattern, which lends support to the additive separability assumption. In addition, allowing for unrestricted match effects between schools and neighborhoods (i.e. a full set of indicator variables for each possible combination of neighborhood and primary/secondary school) only slightly improves the model's fit – e.g. for years of education, the *adjusted* R^2 increases from 0.3710 to 0.3735.

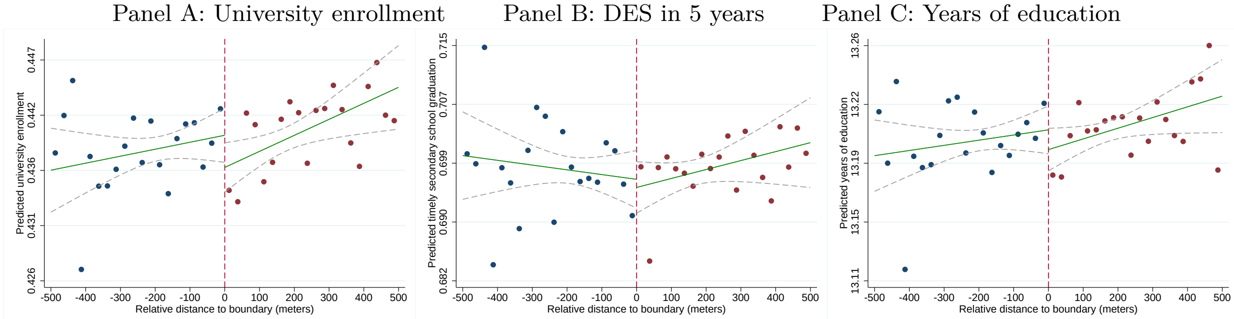
Finally, I collapse the estimated fixed effects, $\hat{\boldsymbol{\delta}}_{s(i)}$ and $\hat{\delta}_n$, at the FSA-level. These averages correspond to $\bar{\Omega}_n^{PR}$ and $\bar{\Lambda}_n^{PR}$. Appendix Figure A9 shows the spatial variation in these measures. The two maps exhibit little overlap – places with low values of the school component $\bar{\Omega}_n^{PR}$ do no necessarily have a low non-school component $\bar{\Lambda}_n^{PR}$.

IV.B Effect of attending better schools

Estimation framework. Here, I present the RD-IV design used to estimate the causal effect of attending better schools on long-term educational outcomes. The approach is based on the fact that schools' catchment areas cut through neighborhoods in such ways that students on opposite sides of a boundary reside in the same communities and enjoy the same neighborhood amenities (Black, 1999). Yet, these boundaries shift the quality of schools two neighbors may be exposed to by varying their default option. I focus on the nearest French primary school boundary from one's residence at baseline throughout.

Figure 2: Balancing Test: Predicted Educational Attainment At Boundaries

Panel A: University enrollment   Panel B: DES in 5 years   Panel C: Years of education



Notes: Predicted educational attainment is given by the fitted values of a regression of the outcome of interest on individual covariates and cohort fixed effects. For each boundary, students assigned the default school with the highest fixed effect $\delta^P_{s(i)}$ are at positive distances. Variables on the vertical axis are first residualized on cohort, boundary and FSA fixed effects. Standard errors are clustered at the boundary level. For visual clarity, students living further than 500 meters away from their nearest boundary are excluded.

leaves unchanged the observation that there is more variance across schools than neighborhoods. If anything, the between-neighborhood variation is noisier. Section V.C presents results of the decomposition exercise using these shrunk estimates.

[30] The correlations are 0.05 and 0.13 for graduating secondary school on time and university enrollment, respectively. The correlations are slightly higher if one uses empirical Bayes shrunk estimates: 0.20, 0.08 and 0.15 for years of education, timely secondary school graduation and university enrollment, respectively.

For each boundary, I first identify which of the two default schools is of better quality, that is which yields greater predicted gains (i.e. has a relatively higher fixed effect $\hat{\delta}_s^P$). Note that because these fixed effects are net of FSA-level variation and secondary school attendance, the "better" school for a given boundary is not necessarily the one where students have the best outcomes in absolute terms. For each student, I define an indicator variable $HighSide_{ib}$ for whether student i resides on the better side of the nearest French primary school boundary b.[31] These indicator variables are used as instruments in the following two-stage regression-discontinuity framework:

$$y_{icnb} = \pi\Omega^{-i}_{s(n(i))} + f(distance_{ib}) + \gamma X_{icnb} + \alpha_b + \alpha_n + \alpha_c + \epsilon_{icnb} \tag{9}$$

$$\Omega^{-i}_{s(n(i))} = \zeta HighSide_b + f(distance_{ib}) + \gamma X_{icnb} + \alpha_b + \alpha_n + \alpha_c + \epsilon_{icnb} \tag{10}$$

where (10) and (9) are first and second stage equations. The dependent variable y_{icnb} is an educational outcome for permanent resident i of neighborhood n. Student-level individual characteristics X_{icnb} are included to improve precision − the point estimates are virtually identical is these covariates are omitted. Each student is matched to the boundary b that is the nearest from her home. The main regressor of interest, $\Omega^{-i}_{s(n(i))}$, is a leave-self-out measure of cumulative school quality over i's entire childhood.[32] In both stages, a control function for distance to the nearest boundary $f(distance_{ib})$ is included, as well as FSA (α_n), boundary (α_b), and cohort (α_c) fixed effects. In the main specification, I follow Lee and Lemieux (2010) and parameterize $f(distance_{ib})$ with a rectangular kernel. Standard errors are clustered at the French primary school boundary level. Robustness of the results to functional form assumptions and bandwidth restrictions is examined in section V.A.

The validity of the RD approach rests on the assumption that right around boundaries, the quality of default school options is as good as random. In education systems where school attendance is fully determined by residence, households may sort right around boundaries, generating discontinuities in sociodemographic characteristics (Bayer, Ferreira and McMillan, 2007). However, in Montreal, opportunities to opt-out of one's default public school and the availability of private schools strongly reduce any incentive to sort at boundaries. For instance, Fack and Grenet (2010) find that the capitalization of school quality in house prices in Paris falls sharply with private school availability, and is effectively zero in areas with many private schools. More importantly, given that rankings of Montreal primary schools are not publicly available, distinguishing good from bad nearby schools is difficult and parents may have little ability to sort at boundaries.[33]

To validate that any jump in school quality at boundaries does not reflect discrete changes in student characteristics, I verify that observable characteristics are balanced around these boundaries. Combining all covariates to generate measures of predicted educational attainment, I find no discontinuity in predicted outcomes. Figure 2 plots predicted outcomes by distance to the nearest boundary, where students assigned to

[31] The boundary-specific higher quality default school is not the one with relatively higher raw outcomes for over a quarter of all permanent residents. In other words, if I were to assign values of $HighSide_{ib}$ on the basis of raw outcomes rather than of adjusted school quality $\hat{\delta}_s^P$, the values of the dummy would flip for a fourth of my sample. Absence of sorting at boundaries on the basis of *adjusted* school quality is consistent with findings that parental preferences are unrelated to school effectiveness once peer quality is accounted for (Abdulkadiroğlu et al., 2017; Rothstein, 2006). Similarly, while school test scores are capitalized in house prices in residence-based school attendance systems, school *value-added* is often not (Imberman and Lovenheim, 2016; Kane, Riegg and Staiger, 2006), with some exceptions (Gibbons, Machin and Silva, 2013).

[32] The childhood school quality measure $\Omega^{-i}_{s(n(i))}$ is obtained by taking the sum of leave-self-out transformations of the primary ($\delta^P_{s(i)}$) and secondary school ($\delta^S_{s(i)}$) fixed effects estimated in Section IV.A. The exact procedure is described in the Data Appendix. Jackknife and split-sample approaches yield almost identical results.

[33] Appendix Figure A10 shows a density plot by distance to boundaries. No excess density is observed on the better side of the threshold (side with relatively better schools in term of university enrollment). A formal McCrary (2008) test finds no statistically significant gap: the log difference in height is 0.006 with a standard error of 0.018.

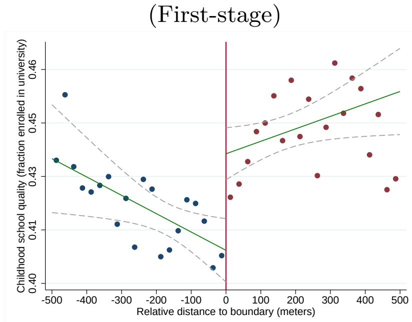
the better school of any boundary-specific pair are depicted on the right of the threshold (positive distances). For visual clarity, I restrict the sample to permanent residents living within 500 meters of their nearest boundary. For each of the three outcomes, no discontinuity in predicted educational attainment can be discerned. The distribution of each covariate taken on its own also appears to be smooth at the threshold (Figures A12, A13 and A14, panels (a) to (j)). Similarly, there is no selective attrition around boundaries (panels (k) and (l)). The associated regression estimates are shown in Table A4.

Note that if non-English families were sorting around boundaries on the basis of their willingness to pay for the quality of French schools, any resulting house price response should lead English families to sort in the other direction, as they all participate in the same housing market. Importantly, there is no discontinuity in the fraction of English families around boundaries (panel (f) of Figures A12, A13 and A14), consistent with the absence of sorting.

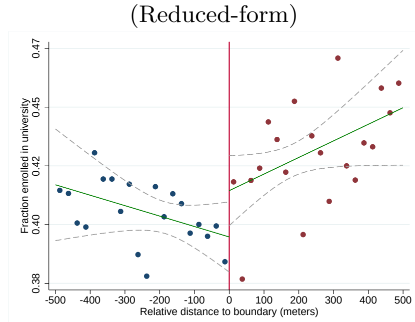
Figure 3: Regression-discontinuity – First-stage and Reduced-form Relationships

All permanent residents

Panel A: Childhood school quality $\Omega^{-i}_{s(n(i))}$ (First-stage)

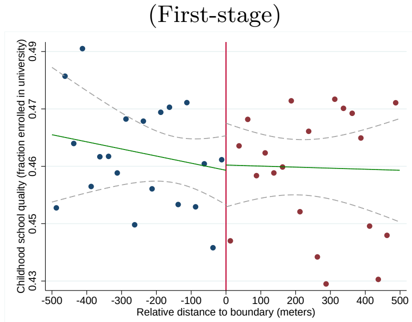


Panel B: University enrollment (Reduced-form)

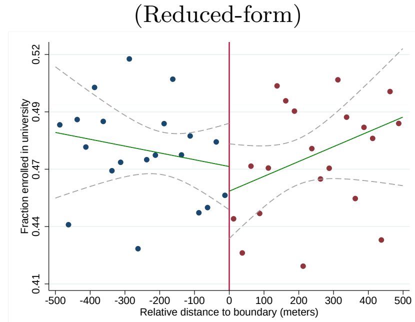


Placebo: Students in English schools only

Panel C: Childhood school quality $\Omega^{-i}_{s(n(i))}$ (First-stage)



Panel D: University enrollment (Reduced-form)



Notes: For each French primary schools boundary, the neighborhood school with greater school quality – in terms of university enrollment – is assigned to the right. The variable on the vertical axis is first residualized on cohort, FSA, and boundary fixed effects. The figure shows the average school quality of schools attended by students at baseline, by distance to the boundary. Attendance recorded at baseline (grade 1). In Panels A and B, the sample includes all permanent residents, and in Panels C and D it is constituted of students enrolled in English schools only. For visual clarity, students living further than 500 meters away from their nearest boundary are excluded.

Results. Figure 3 shows first-stage and reduced-form relationships between distance to boundaries and university enrollment. Panels A and B include all permanent residents, and Panels C and D are placebo tests restricting the sample to students in English schools. The first graph confirms that the instrument has a strong first-stage. Being assigned a better school at baseline does significantly shift the average quality of schools a student attends during childhood ($\Omega^{-i}_{s(n(i))}$), with school quality increasing with distance on both sides of the cutoff. Educational attainment also jumps right at the threshold: students on the better side of a boundary at age 6 are about 2 percentage points more likely to eventually enroll in university later in life. Importantly, there is no break in school quality or university enrollment for students in English schools. The sharp changes observed at the threshold for the full sample are therefore due to schools themselves rather than to some other productive neighborhood characteristic that varies discontinuously and coincides with these boundaries.

Table 2: School Effects – Regression-discontinuity Estimates

	First-stage(s)			Reduced-form	RD-IV
Dependent variable:	Quality of *assigned* school at baseline ($\delta^P_{s(i)}$)	Quality of school *attended* at baseline ($\delta^P_{s(i)}$)	Childhood average school quality ($\Omega^{-i}_{s(n(i))}$)	Outcome	Outcome
	(1)	(2)	(3)	(4)	(5)
Measure of educational attainment					
			All permanent residents		
University enrollment	0.0631	0.0245	0.0328	0.0279	0.8542
	(0.0032)	(0.0027)	(0.0065)	(0.0087)	(0.1645)
Secondary school diploma in 5 years	0.0715	0.0297	0.0337	0.0347	1.0340
	(0.0037)	(0.0025)	(0.0061)	(0.0084)	(0.1618)
Years of schooling	0.2933	0.1157	0.1511	0.1165	0.7739
	(0.0148)	(0.0120)	(0.0298)	(0.0390)	(0.1575)
N	43296	43279	43291	43296	43291
			Placebo: Students in English schools		
University enrollment	0.0632	-0.0017	-0.0012	-0.0098	-
	(0.0044)	(0.0041)	(0.0099)	(0.0157)	-
Secondary school diploma in 5 years	0.0722	0.0031	0.0008	-0.0081	-
	(0.0059)	(0.0026)	(0.0071)	(0.0116)	-
Years of schooling	0.2836	0.0052	0.0104	-0.0448	-
	(0.0204)	(0.0156)	(0.0433)	(0.0655)	-
N	13446	13444	13444	13446	
Cohort fixed effects	x	x	x	x	x
Individual characteristics	x	x	x	x	x
Neighborhood (FSA) fixed effects	x	x	x	x	x
Boundary fixed effects	x	x	x	x	x

Notes: In columns (1) and (2), primary school quality is measured using the fixed effects, $\delta^P_{s(i)}$, estimated in Section IV.A. In column (3), the dependent variable is childhood average school quality $\Omega^{-i}_{s(n(i))}$. Column (5) reports 2SLS estimates of equations (9) and (10). In all specifications, the control function for distance to boundary is linear and allows for different slopes on either side of the threshold. All models include individual characteristics – gender, place of birth indicators, language at home indicators, use of day care, "in difficulty" status at baseline, handicapped status. In the first three rows, the sample includes all permanent residents. In the last three rows, only permanent residents enrolled in English schools are included. All standard errors are clustered at the French primary school boundary level.

Regression results analog to the above figures are presented in Table 2. To increase precision, the main sample imposes no bandwidth restriction and includes all permanent residents. Columns (1) through (4) are first-stage and reduced-form regressions and are estimated by OLS. The average quality gap between French default schools on opposite sides of a shared boundary is 0.063 percentage points (s.e. 0.003) in terms of university enrollment (column (1)). The gap is similar for the subsample of English-school students, indicating that they reside around *French* boundaries that are no different than the boundaries faced by the full sample. For all three main outcomes, differences in default options do translate into significant differences in the quality of schools attended at baseline (gap of 0.025 (s.e. 0.003) in column (2) for university enrollment). Consistent with the visual evidence, this initial shift in default school quality strongly affects average childhood school quality $\Omega^{-i}_{s(n(i))}$ (column (3)). The results in column (4) indicate statistically significant reduced-form relationships between each measure of educational attainment and the assignment variable. For example, students living on the better side of boundary are 0.035 percentage points (s.e. 0.008) more likely to obtain a secondary school diploma in five years than students on the opposite side. Crucially, for columns (2) through (4), all coefficients for placebo tests reported in the bottom panel are close to zero and statistically indistinguishable from zero.

The last column reports two-stage least square estimates of causal school effects. Here, there is some variation across outcomes. The RD-IV coefficient of π is below one for university enrollment (0.854, s.e. 0.165) and years of education (0.774, s.e. 0.158), which implies the presence of some degree of sorting into schools that is not accounted for by place of residence. In contrast, the coefficient for finishing secondary school on time is very close to one. Speculatively, for a given amount of sorting, schools likely have a more direct influence on immediate outcomes such as graduating on time than on higher education investments made later in life.

IV.C Total exposure effects

Estimation framework. The empirical approach used to estimate the combination of school and non-school neighborhood effects – total exposure effects – relies on variation in the timing of moves. I investigate whether the outcomes of movers converge towards those of the permanent residents of the FSA to which they move in proportion with time spent in that neighborhood. As in equation (4), the econometric framework models movers' outcomes as a function of the outcomes of permanent residents of the neighborhoods in which they have resided, weighted by time spent in these locations. The main estimating equation is

$$y_{icmod} = \beta\,(m_i \times \Delta\bar{y}_{od}) + \gamma X_{icmod} + \alpha_{od} + \alpha_m + \alpha_c + \varepsilon_{icmod} \tag{11}$$

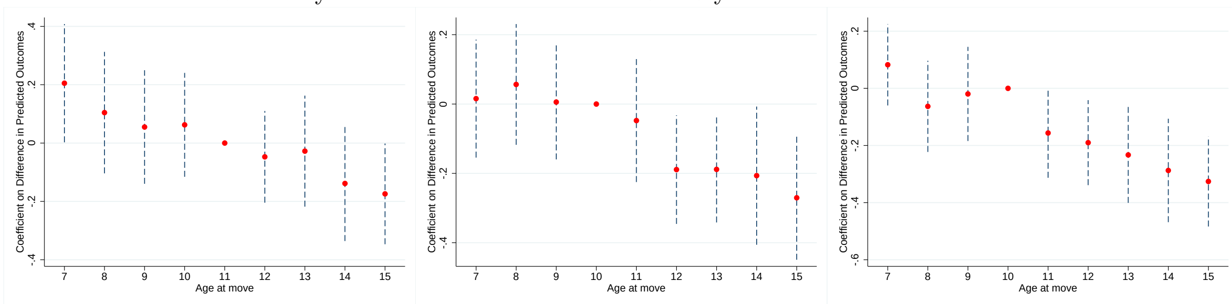
where y_{icod} is some educational outcome of student i, from cohort c, who resided in neighborhood o (origin) at baseline, and moved to neighborhood d (destination) at age m_i. The coefficient of interest β is the annual rate at which outcomes of movers converge to that of the permanent residents in their destination. The origin is the FSA in which students resided at baseline, while the destination is the one in which they lived during the academic year they were aged 15 on September 30. Sorting to better areas is accounted for by origin-by-destination fixed effects (α_{od}) and unobserved differences between students who move at different ages, notably differential disruption costs, are absorbed by age-at-move fixed effects (α_m). Cohorts fixed effects (α_c) are also included to account for the different number of years for which students are tracked in the data. Standard errors are clustered at the destination neighborhood level to allow for arbitrary correlation among families moving to the same place. Note that there is no systematic correlation between m_i and $\Delta\bar{y}_{od}$ in the data. Children who move at early ages are no more likely to move to better or worse areas (relative to their origin) than children who move at later ages (Figure A15). A Kolmogorov-Smirnov test cannot reject the null of equality of distributions of $\Delta\bar{y}_{od}$ between early (age 7-11) and late (age 12-15) movers (p-val=0.22).

To maximize power, in most specifications the sample includes all movers irrespective of the number of times they moved across FSAs, as long as both origin and destination are within Montreal and are not the same. For multiple-times movers, the average quality of neighborhoods exposed to prior to moving to the final destination is therefore measured with error.[34] The model is therefore also estimated on the subsample

[34] To keep the decomposition tractable, I focus on specifications that exploit variation from only two locations (the origin and the destination). Appendix Table A11 reports results for models in which I substitute an exposure-weighted average of neighborhood quality of all locations prior to moving to the final destination for the quality of the origin area $\bar{y}_o$. About $\frac{2}{3}$ of

Figure 4: Semi-parametric Total Exposure Effects

Panel A: University enrollment   Panel B: DES in 5 years   Panel C: Years of education



Notes: Sample includes all movers who remained within Montreal. Observation in FSAs with less than 10 permanent residents are omitted. Coefficients shown are obtained by estimating $y_{icmod} = \sum_{m=7}^{15} \beta_m \left(\Delta \bar{y}_{od} \times 1\{m_i = m\}\right) + \gamma X_{icmod} + \alpha_{od} + \alpha_m + \alpha_c + \epsilon_{icmod}$. Standard errors are clustered at the destination level.

of one-time movers, in which case the econometric model maps directly onto the conceptual framework presented in Section III. In all cases, the sample is always restricted to movers whose origin and destination both have at least 10 permanent residents.

Results. I begin by providing visual evidence of the convergence of movers' outcomes towards those of the permanent residents of their destination by estimating a semi-parametric version of equation (11). More specifically, in Figure 4, I interact $\Delta \bar{y}_{od}$ with a set of indicators for each possible value of age-at-move m_i (age 7 to 15). As expected, the coefficients on $\Delta \bar{y}_{od}$ shrink (increase) with age-at-move (time spent in destination neighborhood). Importantly, they decrease approximately linearly with age-at-move, which validates that the assumption that exposure effects are linear with age is reasonable. The convergence rate is the *slope* of the line that would best fit the coefficients.

Table 3 reports the results for the main outcomes considered – university enrollment, finishing secondary school on time, and years of education. All models are estimated by ordinary least squares and standard errors are clustered at the destination FSA level. For the two binary outcomes, moving one year earlier to a neighborhood where permanent residents exhibit 10-percentage-points higher outcomes, relative to the origin, increases movers' educational attainment by about 0.42 percentage points (column (1)). Extrapolating over 10 years, the cumulative effect would therefore be 4.2 percentage points, or 42% of the difference between permanent residents of the destination and origin locations. These point estimates are all statistically significant at the 1% level. A slightly larger coefficient is obtained for years of education, implying a convergence rate of about 4.9%. Controlling for poor schooling outcomes prior to moving (dummies for the number of times in difficulty, columns (2) and (4)) or restricting the sample to one-time movers (columns (3) and (4)) barely affect the magnitudes of the coefficients.[35] Results for alternative measures of educational attainment mirror the main results (Table A5).[36]

My estimates of total exposure effect are surprisingly close to those reported by Chetty and Hendren (2018*a*), who find a convergence rate of 4% in earnings for millions of movers across commuting zones in the

movers move across FSAs only once, and only 6% move more than three times.

[35] In Table A12, I further include 6-digit postal code fixed effects, thereby restricting the comparison between children who at the age of 15 lived either on the same block or in the same apartment building, as an attempt to absorb as much of the variation in parental income as possible. The fact that many postal codes contain only one observation shrinks the sample size substantially, but the estimated convergence rates remain qualitatively similar to the benchmark estimates.

[36] Allowing permanent residents' outcomes to be cohort-specific ($\bar{y}_{nc}^{PR}$) increases sampling error and therefore yields convergence rates of slightly smaller magnitudes ($3.4 - 4.3\%$). Similar patterns emerge if I use mutually exclusive cohorts to calculate $\bar{y}_n^{PR}$ and to estimate total exposure effects. These results are available upon request.

Table 3: Total Exposure Effects

Sample:	All movers		One-time movers	
	(1)	(2)	(3)	(4)
Measure of educational attainment				
University enrollment	-0.0424	-0.0412	-0.0416	-0.0408
	(0.0090)	(0.0092)	(0.0116)	(0.0115)
Secondary school diploma in 5 years	-0.0421	-0.0402	-0.0506	-0.0502
	(0.0088)	(0.0088)	(0.0117)	(0.0117)
Years of schooling	-0.0488	-0.0471	-0.0444	-0.0435
	(0.0088)	(0.0094)	(0.0103)	(0.0102)
Cohort fixed effects	x	x	x	x
Individual characteristics	x	x	x	x
Age at move fixed effects	x	x	x	x
Origin-by-destination fixed effects	x	x	x	x
Only moved once			x	x
Times in difficulty before moving		x		x
N	24316	24316	15533	15533

Notes: Coefficients shown in the table are convergence rates β. Individual characteristics include gender, immigrant status, allophone status, born in Canada but outside Quebec, English spoken at home, day care use at baseline, "in difficulty" status at baseline, handicapped status. In columns (2) and (4), the model includes a set of dummies for each possible value of number of times in difficulty prior to moving. Standard errors are clustered at the destination neighborhood level. Note that the movers sample contains a total of 25,993 observations, of which 1,677 are singletons and therefore dropped in the estimation.

US. While one may expect a larger influence of neighborhoods at finer levels of geography, my estimates are more likely to be attenuated due to sampling error in the calculation of the average outcomes of permanent residents. Nonetheless, it is remarkable that our findings so closely align given the differences in the locations we study as well as differences in the populations of interest. While movers *across* cities tend to have a slight income advantage relative to stayers, movers *within* Montreal appear to be negatively selected.

The main specification not only assumes that exposure effects are linear with age-at-move, but also that they are linear and symmetric in $\Delta\bar{y}_{od}$. I explore the validity of this assumption by separately estimating the convergence rate for students moving to a better area and for those moving to a worse place. While the point estimates differ for positive and negative moves, I cannot reject that the two coefficients are statistically equal at conventional levels (Table A15, columns (5) and (6)).

Another possible concern is that my preferred definition of neighborhoods might be too large to accurately capture the social context children are exposed to outside of school. Online Appendix E considers census tracts as an alternative unit of analysis to examine whether the results are affected by the choice of geography. Convergence rates are smaller at the census tract level than at the FSA level, possibly because of larger sampling error and of greater sorting of permanent residents at smaller levels of geography. Census tracts may also less precisely reflect all features of the community in which children live and socialize, which is arguably larger than a single census tract. In fact, I find that any benefits of moving to a better census tract are driven by between-FSA moves – no convergence is found for moves that occur within FSAs but across census tracts.

Estimates of the total effect of one year of exposure to a one-unit "better" area are valid under the assumption that the degree of selection to better FSAs does not vary systematically with age. In Section V, a host of robustness checks, including family fixed-effects models, are conducted to corroborate the validity of this assumption.

IV.D Decomposition: Schools or neighborhoods?

Estimation framework. The total convergence rate β reflects the combined effect of changes in school and neighborhood (non-school) quality. To investigate the quantitative importance of schools as a driver of this total effect, I estimate a "horse-race"-type model that simultaneously includes changes in both components of permanent residents' outcomes. Equation (5) suggests the following estimating equation

$$y_{icmod} = \beta_s \left(m_i \times \pi\Delta\Omega_{od}\right) + \beta_n \left(m_i \times \Delta\bar{y}_{od}^{-s}\right) + \gamma X_{icmod} + \alpha_{od} + \alpha_m + \alpha_c + \varepsilon_{icmod} \tag{12}$$

where $\Delta\Omega_{od}$ and $\Delta\bar{y}_{od}^{-s}$ are measured using the fixed effects estimated in Section IV.A. School and non-school marginal effects β_s and β_n are identified from variation in the timing of moves. These are partial regression coefficients that reveal the annual effect of a change in one contextual dimension, holding the other constant. The full convergence rate β is a weighted average of these two marginal effects and can be recovered using the following accounting identity:

$$\beta = \beta_s \left[\frac{Var^r(\pi\Delta\Omega_{od}) + Cov^r(\pi\Delta\Omega_{od}, \Delta\bar{y}_{od}^{-s})}{Var^r(\Delta\bar{y}_{od})}\right] + \beta_n \left[\frac{Var^r(\Delta\bar{y}_{od}^{-s}) + Cov^r(\pi\Delta\Omega_{od}, \Delta\bar{y}_{od}^{-s})}{Var^r(\Delta\bar{y}_{od})}\right] \tag{13}$$

where $Var^r(z)$ and $Cov^r(z)$ respectively denote the variance and covariance of the residuals of $(m_i \times z)$.[37] The total effect of moving to a better area captures independent variation in school and non-school factors (the variances), as well as joint variation in these two dimensions (the covariance). As equation (13) makes clear, because of possible differences in variances, equal effect size (i.e. $\beta_s = \beta_n$) does not imply that schools and other neighborhood factors matter equally. Even if the gains associated with a $\pi\Delta\Omega_{od}$-unit increase in forecast-unbiased school effects are large, schools may nonetheless explain only a small share of the total gains of moving to a better neighborhood if there is little variation in school quality across FSAs (i.e. if $Var^r(\pi\Delta\Omega_{od})$ is small). The empirical counterpart to the school share defined in equation (7) is given by

$$S^{school} = \frac{\beta^{school}}{\beta} = \frac{1}{\beta}\left(\frac{\beta_s Var^r(\pi\Delta\Omega_{od}) + \beta_n Cov^r(\pi\Delta\Omega_{od}, \Delta\bar{y}_{od}^{-s})}{Var^r(\Delta\bar{y}_{od})}\right) \tag{14}$$

and, correspondingly, the fraction of total gains that is not accounted for by causal school effects is

$$S^{non-school} = \frac{\beta^{non-school}}{\beta} = \frac{1}{\beta}\left(\frac{\beta_n Var^r(\Delta\bar{y}_{od}^{-s}) + \beta_s Cov^r(\pi\Delta\Omega_{od}, \Delta\bar{y}_{od}^{-s})}{Var^r(\Delta\bar{y}_{od})}\right). \tag{15}$$

Intuitively, $S^{non-school}$ indicates the fraction of total gains that remains after taking out predicted differences in outcomes between origin and destination on the basis of differences in average forecast-unbiased school effects. In fact, the non-school component $\beta^{non-school}$ defined in equation (15) is numerically equivalent to estimating the benefits of moving to a place that has better outcomes for non-school reasons – i.e. implementing the mover design of equation (11) substituting $\left(m_i \times \Delta\bar{y}_{od}^{-s}\right)$ for $(m_i \times \Delta\bar{y}_{od})$ – and rescaling the resulting coefficient by $\frac{Var^r(\Delta\bar{y}_{od}^{-s})}{Var^r(\Delta\bar{y}_{od})}$ so that it has the same denominator as the full rate.[38]

Results. Again, I start by presenting visual evidence based on semi-parametric estimates. Figure 5 reproduces in light grey the total exposure regression coefficients that were previously shown and displays the

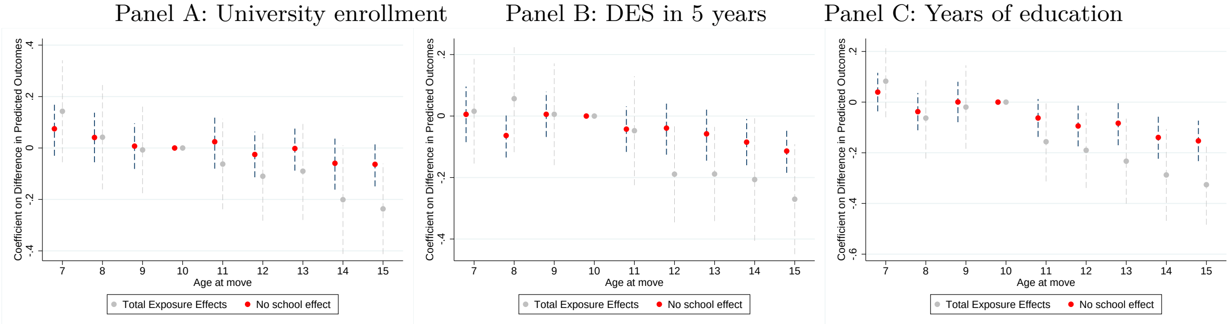
[37] For instance, $Var^r(\Delta\bar{y}_{od})$ is the variance of the residuals of the following regression: $(m_i \times \Delta\bar{y}_{od}) = \gamma X_{icmod} + \alpha_{od} + \alpha_m + \alpha_c + \varepsilon_{icmod}$.

[38] Econometric proofs as well as an interpretation of β_s and β_n in terms of the parameters of the conceptual framework are presented in Online Appendix C.

corresponding $\beta^{non-school}$ components in red.[39] For all three outcomes, the slope of the line that connects these points is considerably flatter, indicating a much lower rate of convergence once school effects have been suppressed.

Figure 5: Semi-parametric Exposure Effects – No School Effect

Panel A: University enrollment Panel B: DES in 5 years Panel C: Years of education



Notes: Sample includes all movers who remained within Montreal. Observation in FSAs with less than 10 permanent residents are omitted. Coefficients in red correspond to age-specific non-school components ($\beta_m^{non-school}$), which are obtained by first estimating a semi-parametric horse-race specification: $y_{icmod} = \sum_{m=7}^{15} \beta_{s,m} \left(\pi \Delta\Omega_{od} \times 1\{m_i = m\}\right) + \sum_{m=7}^{15} \beta_{n,m} \left(\Delta\bar{y}_{od}^{-s} \times 1\{m_i = m\}\right) + \gamma X_{icmod} + \alpha_{od} + \alpha_m + \alpha_c + \varepsilon_{icmod}$. Standard errors are clustered at the destination level and calculated by the delta method.

Estimates of β^{school}, $\beta^{non-school}$ and S^{school} are presented in Table 4. Columns (1) through (3) decompose total exposure effects for all movers, while columns (4) through (6) focuses on one-time movers. In the first column, π is naively set to 1 and school quality is measured by the simple FSA-level average of the sum of primary and secondary school fixed effects ($\bar{\Omega}_n^{PR}$). The non-school component $\beta^{non-school}$ is 0.010 (s.e. 0.004) for university enrollment, 0.012 (s.e. 0.004) for finishing secondary school on time, and 0.014 (s.e. 0.003) for years of education. The corresponding school shares S^{school}, which reflect any variation associated with school attendance, are 76%, 72% and 72%, respectively. In column (2), FSA-level average school quality is measured using permanent residents' leave-self-out childhood school quality $\bar{\Omega}_{s(n)}^{-i}$, the variable used to estimate π. This minor change in measurement has little effect on the results – the fraction of total exposure effects explained by schools remains in the vicinity of $70 - 75\%$.

Decompositions based on biased measures of school quality may overstate the importance of schools. My preferred specification, in column (3), uses forecast-unbiased school quality $\pi\bar{\Omega}_{s(n)}^{-i}$. Here, the school share drops to 65% for university enrollment, to 71% for finishing secondary school on time, and to 54% for years of education. The pattern is similar for one-time movers, with school shares ranging between 55% and 76% in column (6). The values of β^{school} reported in column (6) suggest that movers' outcomes would converge at rates of 2.7% (s.e. 0.008), 3.9% (s.e. 0.011) and 2.5% (s.e. 0.007) towards mean outcomes of permanent residents for university enrollment, timely secondary school graduation, and years of education, respectively, on the basis of average differences in causal school effects alone.

In Online Appendix E, the decomposition exercise is implemented using census tracts instead of FSAs. The results are very similar to analyses performed at the FSA level, with school shares S^{school} ranging from 53% for university enrollment to 76% for timely graduation from secondary school.

For completeness, I also consider an alternative definition of the non-school share which measures the fraction of β that would remain if movers were not to benefit from moving to places with better schools, which is obtained by setting the marginal effect of geographic differences in school effects on movers, β_s, to zero in equation (13). This approach differs from the main decomposition in equation (14) only in how it allocates the covariance terms between the school and non-school components. Results are shown in Table

[39] Appendix Figure A16 plots estimates of $\beta_{s,m}$ and $\beta_{n,m}$, which are roughly linear with age-at-move.

Table 4: Decomposition of Total Exposure Effects

Sample:	All movers			One-time movers		
	(1)	(2)	(3)	(4)	(5)	(6)
University enrollment						
			Total exposure effects			
β	-0.0424	-0.0424	-0.0424	-0.0416	-0.0416	-0.0416
	(0.0090)	(0.0090)	(0.0090)	(0.0116)	(0.0116)	(0.0116)
			School and non-school components			
β^{school}	-0.0321	-0.0320	-0.0274	-0.0320	-0.0318	-0.0272
	(0.0071)	(0.0070)	(0.0060)	(0.0092)	(0.0090)	(0.0077)
$\beta^{non\text{-}school}$	-0.0103	-0.0104	-0.0150	-0.0096	-0.0097	-0.0144
	(0.0038)	(0.0038)	(0.0043)	(0.0058)	(0.0059)	(0.0063)
Share school effects (S^{school})	76%	76%	65%	77%	77%	65%
	(0.0678)	(0.0678)	(0.0579)	(0.1125)	(0.1124)	(0.0960)
Secondary school diploma in 5 years						
			Total exposure effects			
β	-0.0421	-0.0421	-0.0421	-0.0506	-0.0506	-0.0506
	(0.0088)	(0.0088)	(0.0088)	(0.0117)	(0.0117)	(0.0117)
			School and non-school components			
β^{school}	-0.0305	-0.0290	-0.0299	-0.0398	-0.0373	-0.0386
	(0.0083)	(0.0079)	(0.0082)	(0.0106)	(0.0104)	(0.0108)
$\beta^{non\text{-}school}$	-0.0116	-0.0131	-0.0121	-0.0108	-0.0133	-0.0120
	(0.0039)	(0.0041)	(0.0041)	(0.0049)	(0.0050)	(0.0050)
Share school effects (S^{school})	72%	69%	71%	79%	74%	76%
	(0.0918)	(0.0905)	(0.0936)	(0.0889)	(0.0890)	(0.0920)
Years of education						
			Total exposure effects			
β	-0.0488	-0.0488	-0.0488	-0.0444	-0.0444	-0.0444
	(0.0088)	(0.0088)	(0.0088)	(0.0103)	(0.0103)	(0.0103)
			School and non-school components			
β^{school}	-0.0350	-0.0340	-0.0263	-0.0328	-0.0316	-0.0245
	(0.0075)	(0.0072)	(0.0055)	(0.0088)	(0.0085)	(0.0065)
$\beta^{non\text{-}school}$	-0.0138	-0.0148	-0.0225	-0.0115	-0.0127	-0.0199
	(0.0034)	(0.0035)	(0.0042)	(0.0045)	(0.0046)	(0.0053)
Share school effects (S^{school})	72%	70%	54%	74%	71%	55%
	(0.0597)	(0.0570)	(0.0441)	(0.0875)	(0.0851)	(0.0659)
Measure of school quality	$\pi\Omega_{s(n)}$	$\pi\Omega^{-i}_{s(n)}$	$\pi\Omega^{-i}_{s(n)}$	$\pi\Omega_{s(n)}$	$\pi\Omega^{-i}_{s(n)}$	$\pi\Omega^{-i}_{s(n)}$
π	1	1	RD estimate	1	1	RD estimate

Notes: Sample restricted to movers within Montreal. Standard errors are clustered at the destination FSA level, and obtained by the delta method. Parameters β^{school}, $\beta^{non-school}$ and S^{school} are calculated using equations (12), (13) and (14). In columns (1) and (2), π is set to one. In column (3), π is given by the RD-IV estimates reported in Table 2. In columns (1) through (3), total exposure effects β correspond to estimates reported in Table 3, column (1). In columns (4) through (6), β correspond to estimates reported in Table 3, column (3).

A6. Because the covariance term $Cov^r(\pi\Delta\Omega_{od}, \Delta\bar{y}_{od}^{-s})$ is considerably smaller than the two variances, this alternative approach yields results that are fairly similar to the main approach. In specifications that adjusts for π, the school share range from 46% to 72% for all movers and from 52% to 76% for one-time movers alone.

Overall, the decomposition analysis indicates that schools matter more than neighborhoods for long-term educational attainment, with most estimates that account for bias in measures of school quality ranging between 50% and 70%. Nonetheless, schools do not *fully* account for total exposure effects – neighborhoods do have a small independent effect on human capital accumulation. Note that the school contribution to total exposure effects may be larger in contexts where school choice is more restricted, since the school share S^{school} increases with the compliance factor $c_{i(o,d)}$.

V Robustness

V.A Regression-discontinuity estimates

The benchmark specification for estimating school effects imposes several restrictions. Firstly, it assumes that the relationship between distance to the boundary and student outcomes is linear. Appendix Table A7 allows for a quadratic functional form. RD-IV estimates for finishing secondary school on time and years of education appear insensitive to this assumption. The estimate of π with quadratic functions for university enrollment (0.71), however, is smaller than the baseline (0.85). Using a triangular kernel for the control function yields results almost identical to the baseline (Appendix Table A8).

Sensitivity to bandwidth restrictions is examined in Appendix Figure A17. Moving along the horizontal axis, I gradually expand the sample by including students living farther away from boundaries. The point estimates fluctuate across sample restrictions, following no monotonic pattern. For instance, for university enrollment, keeping only students living within 750 meters of a boundary yields a relatively small π coefficient of 0.62, while further restricting the bandwidth to 300 meters produces a coefficient very close to the baseline (0.83). These movements in point estimates are plausibly driven by differences across the set of schools and neighborhoods that are dropped when the bandwidth is changed. For instance, denser parts of Montreal are unaffected by these restrictions since students living in these areas all live very close to a boundary. Large distances from boundaries are only observed in the suburbs – about half my sample of permanent residents live within 200 meters of their nearest boundary. In addition, Table A9 presents results using optimal bandwidths based on the methods developed in Calonico, Cattaneo and Titiunik (2014), which vary between 300 and 400 meters, depending on the outcome. Overall, most estimates shown in Figure A17 and in Table A9 remain within short range of the baseline results. The conclusions of the decomposition analysis are therefore unaffected by the choice of bandwidth – the vast majority of the associated estimates of the school share S^{school} fall between 50% and 70% (Appendix Figure A18).

A separate issue arise in the decomposition exercise: the RD estimates may reflect local average treatment effects for a different subpopulation than the one that identifies total exposure effects. To address concerns related to heterogeneous treatment effects, I use a nearest-neighbor matching algorithm to re-weight the sample of permanent residents so that their distribution of observables matches the one of the movers' subsample (Jann, 2017; Abadie and Imbens, 2011). School effects estimates for this reweighted sample are close to the baseline results (Table A10).

Finally, in Online Appendix D, I conduct further specifications checks on school effects. These additional analyses include testing for locally constant effects around the cutoff (Dong and Lewbel, 2015), verifying that estimates of π are stable across boundaries that generate gaps in the quality of default options of different magnitudes, and calculating model-based estimates of π using the movers design rather than the RD approach. All of the associated results are in line with the range of estimates of π reported in the main text.

V.B Movers estimates

My estimates of total exposure effects may be biased if students with higher (lower) unobserved family inputs who move to better (worse) areas tend to do so earlier. This section examines the validity of the identifying assumption by running a set of robustness checks that address issues of time-invariant and time-varying unobserved heterogeneity.

Table 5: Total Exposure Effects – Siblings Subsample

Sample:	Siblings only			
	(1)	(2)	(3)	(4)
Measure of educational attainment				
University enrollment	-0.0453	-0.0571	-0.0478	-0.0504
	(0.0365)	(0.0359)	(0.0275)	(0.0274)
Secondary school diploma in 5 years	-0.0242	-0.0434	-0.0392	-0.05
	(0.0378)	(0.0365)	(0.0301)	(0.0300)
Years of schooling	-0.0453	-0.0629	-0.0444	-0.0486
	(0.0311)	(0.0299)	(0.0236)	(0.0233)
Cohort fixed effects	x	x	x	x
Individual characteristics	x	x	x	x
Age at move fixed effects	x	x	x	x
Origin-by-destination fixed effects	x	x		
Household fixed effects			x	x
Times in difficulty before moving		x		x
N	3674	3674	3674	3674

Notes: Restricted to households in which siblings lived at the same address for at least 75% of the observed years. In columns (2) and (4), the model includes a set of dummies for each possible value of number of times in difficulty prior to moving. Standard errors are clustered at the household level.

Within-family exposure effects. The first test I run involves estimating the exposure model with household fixed effects to account for any time-invariant family unobserved heterogeneity. In this specification, identification relies on age differences between siblings. In this context, positive exposure effects would generate a relationship between the change in neighborhood and school quality, on one hand, and the difference in educational outcomes among siblings, on the other hand, that varies proportionally to the age-difference of siblings.

Since siblings are not directly identified in the data, I match students using unique moves at a very fine level of geography. More precisely, I assume that two students who move from and to the exact same six-digit postal codes in the same year must belong to the same household.[40] Many household units are not consistent over time given the prevalence of step- and blended-families. For instance, two students from different biological parents may have been living under the same roof only for a fraction of their lives. I therefore exclude household units for which the children have lived at a common postal code for less than 75% of the years for which I can observe them.

In columns (1) and (2) of Table 5, I estimate the exposure model with origin-by-destination fixed effects on the subsample of siblings. Standard errors are considerably larger than in the main specification since the sample size is much smaller, but the point estimates are in line with the main results. In columns (3) and (4), I substitute family fixed effects for the origin-by-destination fixed effects to account for any time-invariant heterogeneity across families and still find convergence rates of about 4.5%. These results support the idea that the estimated exposure effects are not driven by differences in unobservable time-invariant family characteristics.

Balance of observables. The second approach tests for balance of covariates to verify that variation in the interaction term is arguably random conditional on age-at-move and origin-by-destination fixed effects.

[40] Out of the original 100,929 students, this method identifies about 13,000 siblings attached to roughly 6,000 different households.

Here, I estimate the exposure model using individual characteristics as dependent variables. Pei, Pischke and Schwandt (2017) show that putting the covariates on the left-hand side is a more powerful test than gradually adding or removing these variables from the right-hand side of the main estimating equation, particularly if the individual characteristics are poor measures of the underlying confounders they are meant to account for (e.g. being "in difficulty" is certainly a noisy measure of academic abilities).

Results of the test are shown in Table A13. In columns (1) and (2), I use years of education of permanent residents to measure $\Delta\bar{y}_{od}$. Finishing secondary school on time and university enrollment are used in columns (3) and (4) and columns (5) and (6), respectively. The coefficients on immigrant status are marginally significant at the 5% level for some, but not all outcomes. In Montreal, immigrants do obtain more post-secondary education than domestic students. It might also be the case that they tend to move later given that they may have less prior information about neighborhoods than native-born parents. Overall, most coefficients in the table are very small and statistically indistinguishable from zero. As a result, the baseline convergence rate is materially unchanged whether covariates are included or not, despite the fact that these covariates have non-trivial explanatory power with respect to educational attainment – e.g. for years of education, their inclusion increases the adjusted R^2 from 0.188 to 0.307.

Selection on time-varying observables. Another possible source of concern is that length of exposure to a one-unit better area mirrors exposure to different family circumstances. One may be worried that if a move is triggered by a change in marital status or income, age-specific unobserved parental inputs may have also changed in proportion with m_i. Unfortunately, my data set includes very few time-varying individual characteristics. For instance, parental income and marital status of parents are not observed. To account for possible changes in family circumstances that coincides with a move, I instead control for differences in census tract characteristics between the areas in which student i resides after and before the move, as well as the interaction of these differences with age-at-move. Because census tracts are considerably smaller than FSAs, these controls vary within origin-by-destination cells and so the main effects are identified. The characteristics I consider are median household income, average dwelling value, percentage low income, percentage of adults with some college education, and fraction of lone parent families. Those are all obtained from the 2001 Canadian Census. The inclusion of these variables accounts for changes in family circumstances that are correlated with changes in neighborhood attributes, as well as any sorting on the basis of these observable neighborhood characteristics. Similarly, Altonji and Mansfield (2014) control for group-level average individual characteristics – arguably the basis on which households sort into neighborhoods – to account for unobserved individual heterogeneity, and thereby obtain a lower bound on contextual effects. For example, a positive income shock may be associated with both a move to an area where property value is higher than in the origin and an increase in parental inputs. For any unobserved variable to generate bias in the exposure estimates under this specification, the confounding variable would have to generate variation orthogonal to changes in these neighborhood attributes. The inclusion of these neighborhood attributes likely absorbs part of the causal exposure effect of interest, and therefore over-adjusts for changes in family circumstances.

Results are quite robust to the inclusion of these controls (Table A14). In columns (1) through (5), I control for changes in one time-varying characteristic at a time. In all of these cases, the exposure effects remain stable around 4 to 4.5%. Among all considered variables, the local fraction of lone-parent families is the one that most affects the main exposure effects. Yet, even in this case, the exposure effects remain large (≈4%). In column (6), I include all controls simultaneously – exposure effects fall just under 4% and remain strongly statistically significant.

Event-study. Next, I investigate whether students who move to better areas exhibit different trends in learning difficulties prior to moving. The idea is that family circumstances plausibly directly affect the likelihood that a student struggles in school, hence changes in unobserved family inputs should be reflected in the probability of being identified in difficulty. I leverage year-to-year variation in $Diff_{iod,t}$, the indicator of whether student i was in difficulty in year t, to create an index of relative learning difficulties that summarizes the way movers compare to permanent residents in their origin and destination $\sigma_{od(i,t)} = \frac{Diff_{iod,t}-\overline{Diff}_{o,t}}{\overline{Diff}_{d,t}-\overline{Diff}_{o,t}}$ where $\overline{Diff}_{n,t}$ is the fraction of permanent residents of FSA n that were in difficulty at time t (years since started grade 1). This index takes a value of zero if mover i's difficulty status is the same as the average in her origin and a value of one if it is equal to the average in the destination. An increase in $\sigma_{od(i,t)}$ over time indicates that student i's success in school (or lack of thereof) converges towards that of permanent residents in the destination relative to the origin.

I investigate patterns in $\sigma_{od(i,t)}$ around the time of moves. For instance, a positive pre-trend would indicate that movers started converging towards permanent residents of the destination before they even moved. Such a pattern could arise if moves to certain areas occur as a result of gradual changes in family circumstances. For example, if divorces are preceded by an erosion of the quality of the parents' relationship *and* are disproportionately followed by moves to worse places, my estimates of total exposure effects could be biased.

Figure A19 shows an event-study analysis, in which $\sigma_{od(i,t)}$ is regressed on time dummies (relative to year of move) and where observations are weighted by $\left(\overline{Diff}_{d,t} - \overline{Diff}_{o,t}\right)^2$ as in Bronnenberg, Dubé and Gentzkow (2012). A jump in $\sigma_{od(i,t)}$ occurs on impact, and students schooling difficulties then converge gradually towards the destination's average (Panel A). Importantly, there is no discernible pre-trend – coefficients are stable prior to moving. These results are consistent with Aaronson (1998), who finds no systematic pattern between pre-move changes in family circumstances and the quality of the destination neighborhood quality. For instance, moves preceded by a divorce are just as likely to lead to a better than to a worse destination.

Because students move at different ages, the panel is not balanced. As a result, any pre- and post-move trend may be the result of changes in the sample's composition. In Panel B, I follow Finkelstein, Gentzkow and Williams (2016) and include student fixed effects to address this issue. While the post-move trend now disappears, the jump at the time of the move remains significant. One concern is that this sudden jump is the product of a sharp and sudden change in family inputs. In Panel C and D, I therefore distinguish between students who did and did not switch school the year they moved. Students who did not change school at the time of the move show no jump in $\sigma_{od(i,t)}$, suggesting that the break reflects a change in the schooling environment (e.g. differences in schools' propensity to flag marginal students) rather than a change in family inputs. Overall, I find no evidence of pre-trends in schooling difficulties.

V.C Decomposition

Movers' school attendance. The main decomposition approach projects differences in the quality of schools attended by permanent residents onto movers. One concern is that measures of school quality estimated on a sample of non-movers may not accurately capture the true change in school quality experienced by movers. In this subsection, I instead directly account for movers' school attendance in the baseline total exposure effect model The estimating equation becomes

$$y_{icmod} = \beta\left(m_i \times \Delta\bar{y}_{od}\right) + \alpha_{s(0)} + \alpha_{s(A)} + \gamma X_{icod} + \alpha_{od} + \alpha_m + \alpha_c + \epsilon_{icmod} \quad (16)$$

where $\alpha_{s(0)}$ and $\alpha_{s(A)}$ are sets of fixed effects for schools attended at baseline and at age 15, respectively. To account for variation in length of exposure to better schools that may be correlated with neighborhood exposure, the school fixed effects are allowed to vary linearly with age-at-move, i.e. $\alpha_{s(a)} = \alpha^0_{s(a)} + \alpha^1_{s(a)} \times m_i$, which is equivalent to allowing age-at-move effects to have a different slope in each school.

Table 6: Exposure Effects Net of Movers' School Attendance

	Convergence rate			Decomposition: $\beta^{non\text{-}school}$ for π=1
	(1)	(2)	(3)	(4)
University enrollment	-0.0424	-0.0111	-0.0123	-0.0103
	(0.0090)	(0.0101)	(0.0089)	(0.0038)
Secondary school diploma in 5 years	-0.0421	-0.0117	-0.0075	-0.0116
	(0.0088)	(0.0093)	(0.0081)	(0.0039)
Years of education	-0.0488	-0.0178	-0.0181	-0.0138
	(0.0088)	(0.0079)	(0.0075)	(0.0034)
Cohort fixed effects	x	x	x	x
Individual characteristics	x	x	x	x
Age at move fixed effects	x	x	x	x
Origin-by-destination fixed effects	x	x	x	x
School fixed effects				
(o) School at baseline		x	x	
(o) School at baseline * age-at-move (linear)		x		
(o) School at baseline * years-exposure			x	
(d) School at age 15		x	x	
(d) School at age 15 * age-at-move (linear)		x		
(d) School at age 15 * years-exposure			x	
N	24316	24244	24244	24316

Notes: Primary school fixed effects are based on school attendance at baseline. Secondary school fixed effects are based on school attendance at age 15. In columns (2) and (3), school fixed effects are linearly interacted with age-at-move and years of exposure, respectively. Standard errors are clustered at the destination neighborhood level.

The results are presented in Table 6. Benchmark estimates of total exposure effects are reproduced in column (1). In column (2), fixed effects for schools attended at the beginning (the "origin" school) and at the end (the "destination" school) of the exposure period, as well as interactions with age-at-move, are added. Net of school attendance, the annual exposure effects shrink substantially to 1.1% for university enrollment, 1.2% for completing secondary school on time, and 1.8% for years of education. As a further robustness check, in column (3) the school fixed effects are interacted with the actual number of years spent in the associated schools instead of with age-at-move.

To get a sense of how these results compare to the ones based on the main decomposition approach, I report the corresponding estimates of $\beta^{non-school}$ in column (4). Because the inclusion of school fixed effects in equation (16) absorbs any variation associated with school attendance (i.e. it may account for more than just the causal component), I report unadjusted estimates of $\beta^{non-school}$ for which $\pi = 1$ to make the two sets of estimates comparable. Looking across columns (2) and (4), one find the point estimates to be strikingly close (e.g. 1.1% vs 1.0% for university enrollment), suggesting that $\Delta\Omega_{od}$ does a good job of representing the effective change in school quality faced by movers.

Sampling error. One possibility is that sampling error affects the estimation of school and neighborhood fixed effects differentially. For instance, it might be that estimation error accounts for a larger fraction of the variance in $\Delta\Omega_{od}$ than of the variance in $\Delta\Lambda_{od}$. To verify that this is not the case, I use empirical Bayes techniques (Chandra et al., 2016; Best, Hjort and Szakonyi, 2017; Kane and Staiger, 2008) to shrink towards zero the school and neighborhood fixed effects obtained from equation (8). The associated decomposition results are shown in Table A16.

The first row shows "one-step" total convergence rates estimated using shrunk estimates of permanent residents' mean outcomes, $\bar{y}_n^{EB}$, in equation (11). These rates are slightly larger than the ones presented in Table 3, varying between 4.5% and 5%, which imply that my main estimates suffer from a small attenuation bias. The second row reports "two-step" convergence rates obtained using shrunk estimates of school and neighborhood fixed effects, $\Omega_{s(n(i))}^{EB}$ and Λ_n^{EB}, to estimate equation (12) and recover the full rate using equation (13). The one-step and two-step rates differ because the sum of two shrunk estimates is not the same as shrunk estimates of the sum. In other words, while $\bar{y}_n = \Lambda_n + \Omega_n$ holds true by construction, the equality does not hold for shrunk estimates, $\bar{y}_n^{EB} \neq \Lambda_n^{EB} + \Omega_n^{EB}$. However, I cannot reject the null that the one-step and two-step estimates are equal.

The school share is calculated using the two-step total convergence rate as the proper denominator to ensure that the school and non-school share sum to one. The shares of total exposure effects due to schools reported in this table do not account for the endogeneity of school attendance and should therefore be compared to the corresponding results presented in column (1) of Table 4. For all three outcomes, adjusting for measurement error only reinforces the conclusion that school effects account for most of the benefits of moving to a better area, with school shares exceeding 80% when using empirical Bayes shrunk estimates.

VI Conclusion

Establishing whether schools drive neighborhood exposure effects is crucial on a policy level to inform the development of community-wide versus in-school intervention programs. Yet, isolating the effects of neighborhoods from those of schools is a difficult task since in most places students are allocated to schools on the basis of residence. This paper overcomes these difficulties by bringing together two research designs in order to isolate the fraction of total long-term exposure effects that is driven by school effects.

The first contribution of this paper is to break the mechanical link between the two dimensions by exploiting institutional features of Quebec's education system. In Montreal, default options influence parents' decision over which schools their child will attend. Building upon this observation, I find that the quality of the primary and secondary schools children attend have large effects on their educational attainment. More precisely, immediate neighbors living on opposite sides of a French primary school boundary at age 6 exhibit significantly different propensities to enroll into university more than 10 years later.

My second set of results demonstrates that children who move to a better neighborhood at a young age benefit substantially from this change. In particular, I successfully replicate the findings of linear exposure effects of Chetty and Hendren (2018*a*) using within-city variation and implementing their methods in a different setting, looking at a much smaller scale of geography and examining different outcomes. My estimates suggest that movers' educational attainment improve linearly with each year spent in a better location at an annual rate of approximately 4.5%.

The main result of the paper is that schools are the main driver of total childhood exposure effects. Decompositions that take into account the endogeneity of school quality indicate that between 50% and

70% of the educational benefits of moving to a better location are due to schools rather than neighborhoods themselves. These findings strongly corroborate earlier conclusions made on the relative importance of schools and neighborhoods (Dobbie and Fryer, 2015; Fryer and Katz, 2013; Oreopoulos, 2012). By showing that spatial inequalities in long-term educational attainment are partly rooted in the quality of schools children attend, the results bear important policy implications. They notably suggest that school reforms or interventions might be more effective than community programs or relocation policies in raising educational attainment. However, the policy implications one can draw from the results presented in this paper are limited by the fact that the way school quality is measured conflates many different inputs, such as peer, teacher as well as school principal quality.

While the magnitude of the estimated exposure effects, and the main conclusion of this paper more broadly, may reflect a social reality unique to Montreal, I believe the results are very informative for other contexts as well. For instance, because of Quebec's open enrollment policy and the unusual availability of private school options in Montreal, the link between school attendance and residence is relatively loose. Hence, in jurisdictions where schools and neighborhoods are tightly linked, one may expect schools to contribute even more to spatial inequalities in educational attainment than the results in this paper suggest. I leave for future research the question of whether this conclusion extends to other socio-economic outcomes such as earnings and criminal behavior.

References

Aaronson, Daniel. 1998. "Using sibling data to estimate the impact of neighborhoods on children's educational outcomes." *Journal of Human Resources*, 915–946.

Abadie, Alberto, and Guido W Imbens. 2011. "Bias-corrected matching estimators for average treatment effects." *Journal of Business & Economic Statistics*, 29(1): 1–11.

Abdulkadiroğlu, Atila, Joshua Angrist, and Parag Pathak. 2014. "The elite illusion: Achievement effects at Boston and New York exam schools." *Econometrica*, 82(1): 137–196.

Abdulkadiroğlu, Atila, Joshua D. Angrist, Susan M. Dynarski, Thomas J. Kane, and Parag A. Pathak. 2011. "Accountability and flexibility in public schools: Evidence from Boston's charters and pilots." *The Quarterly Journal of Economics*, 126(2): 699–748.

Abdulkadiroğlu, Atila, Parag A. Pathak, Jonathan Schellenberg, and Christopher R. Walters. 2017. "Do Parents Value School Effectiveness?" National Bureau of Economic Research, Inc NBER Working Papers 23912.

Abowd, John M, Francis Kramarz, and David N Margolis. 1999. "High wage workers and high wage firms." *Econometrica*, 67(2): 251–333.

Altonji, Joseph G., and Richard K. Mansfield. 2014. "Group-Average Observables as Controls for Sorting on Unobservables When Estimating Group Treatment Effects: the Case of School and Neighborhood Effects." National Bureau of Economic Research, Inc NBER Working Papers 20781.

Angrist, Joshua D., and Alan B. Krueger. 1991. "Does compulsory school attendance affect schooling and earnings?" *The Quarterly Journal of Economics*, 106(4): 979–1014.

Angrist, Joshua D., Peter D. Hull, Parag A. Pathak, and Christopher R. Walters. 2017. "Leveraging lotteries for school value-added: Testing and estimation." *The Quarterly Journal of Economics*, 132(2): 871–919.

Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman. 2016. "School Quality and the Gender Gap in Educational Achievement." *American Economic Review*, 106(5): 289–295.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2007. "A unified framework for measuring preferences for schools and neighborhoods." *Journal of political economy*, 115(4): 588–638.

Best, Michael Carlos, Jonas Hjort, and David Szakonyi. 2017. "Individuals and Organizations as Sources of State Effectiveness, and Consequences for Policy." National Bureau of Economic Research 23350.

Billings, Stephen B., David J. Deming, and Stephen L. Ross. 2016. "Partners in Crime: Schools, Neighborhoods and the Formation of Criminal Networks." National Bureau of Economic Research, Inc NBER Working Papers 21962.

Black, Sandra E. 1999. "Do better schools matter? Parental valuation of elementary education." *Quarterly journal of economics*, 577–599.

Boudarbat, Brahim, Thomas Lemieux, and W. Craig Riddell. 2010. "The Evolution of the Returns to Human Capital in Canada, 1980-2005." *Canadian Public Policy*, 36(1): 63–89.

Bronnenberg, Bart J., Jean-Pierre H. Dubé, and Matthew Gentzkow. 2012. "The evolution of brand preferences: Evidence from consumer migration." *The American Economic Review*, 102(6): 2472–2508.

Burdick-Will, Julia, Jens Ludwig, Stephen W. Raudenbush, Robert J. Sampson, Lisa Sanbonmatsu, and Patrick Sharkey. 2011. "Converging evidence for neighborhood effects on children's test scores: An experimental, quasi-experimental, and observational comparison." In *Whither opportunity?: Rising inequality, schools, and children's life chances.* , ed. Greg J. Duncan and Richard J. Murnane, Chapter 12. Russell Sage Foundation.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik. 2014. "Robust nonparametric confidence intervals for regression-discontinuity designs." *Econometrica*, 82(6): 2295–2326.

Card, David. 2001. "Estimating the return to schooling: Progress on some persistent econometric problems." *Econometrica*, 69(5): 1127–1160.

Card, David, and Jesse Rothstein. 2007. "Racial segregation and the black-white test score gap." *Journal of Public Economics*, 91(11-12): 2158–2184.

Card, David, Ciprian Domnisoru, and Lowell Taylor. 2018. "The Intergenerational Transmission of Human Capital: Evidence from the Golden Era of Upward Mobility."

Card, David, Jorg Heining, and Patrick Kline. 2013. "Workplace Heterogeneity and the Rise of West German Wage Inequality." *The Quarterly Journal of Economics*, 128(3): 967–1015.

Card, David, Martin D. Dooley, and A. Abigail Payne. 2010. "School competition and efficiency with publicly funded Catholic schools." *American Economic Journal: Applied Economics*, 2(4): 150–176.

Carlson, Deven, and Joshua M. Cowen. 2015. "Student neighborhoods, schools, and test score growth: Evidence from Milwaukee, Wisconsin." *Sociology of Education*, 88(1): 38–55.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2016. "Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector." *The American Economic Review*, 106(8): 2110–2144.

Chetty, Raj. 2015. "Behavioral economics and public policy: A pragmatic perspective." *The American Economic Review*, 105(5): 1–33.

Chetty, Raj, and Nathaniel Hendren. 2018*a*. "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects." *The Quarterly Journal of Economics*, forthcoming.

Chetty, Raj, and Nathaniel Hendren. 2018*b*. "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates." *The Quarterly Journal of Economics*, forthcoming.

Chetty, Raj, John N. Friedman, and Emmanuel Saez. 2013. "Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings." *The American Economic Review*, 103(7): 2683–2721.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014*a*. "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates." *American Economic Review*, 104(9): 2593–2632.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014*b*. "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood." *The American Economic Review*, 104(9): 2633–2679.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review*, 106(4): 855–902.

Chyn, Eric. 2016. "Moved to Opportunity: The Long-Run Effect of Public Housing Demolition on Labor Market Outcomes of Children."

Cullen, Julie Berry, Brian A. Jacob, and Steven Levitt. 2006. "The effect of school choice on participants: Evidence from randomized lotteries." *Econometrica*, 74(5): 1191–1230.

Deming, David J. 2014. "Using school choice lotteries to test measures of school effectiveness." *The American Economic Review*, 104(5): 406–411.

Deming, David J., Justine S. Hastings, Thomas J. Kane, and Douglas O. Staiger. 2014. "School choice, school quality, and postsecondary attainment." *The American Economic Review*, 104(3): 991–1013.

Dobbie, Will, and Roland G. Fryer. 2011. "Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone." *American Economic Journal: Applied Economics*, 3(3): 158–187.

Dobbie, Will, and Roland G. Fryer. 2013. "Getting beneath the veil of effective schools: Evidence from New York City." *American Economic Journal: Applied Economics*, 5(4): 28–60.

Dobbie, Will, and Roland G. Fryer. 2015. "The medium-term impacts of high-achieving charter schools." *Journal of Political Economy*, 123(5): 985–1037.

Dong, Yingying, and Arthur Lewbel. 2015. "Identifying the effect of changing the policy threshold in regression discontinuity models." *Review of Economics and Statistics*, 97(5): 1081–1092.

Duhaime-Ross, Alix. 2015. “Three essays in the economics of education: evidence from Canadian policies.” PhD diss. University of British Columbia.

Fack, Gabrielle, and Julien Grenet. 2010. “When do better schools raise housing prices? Evidence from Paris public and private schools.” *Journal of public Economics*, 94(1): 59–77.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2016. “Sources of Geographic Variation in Health Care: Evidence From Patient Migration.” *The Quarterly Journal of Economics*, 131(4): 1681–1726.

Fryer, Roland G., and Lawrence F. Katz. 2013. “Achieving Escape Velocity: Neighborhood and School Interventions to Reduce Persistent Inequality.” *American Economic Review*, 103(3): 232–37.

Gibbons, Stephen, Olmo Silva, and Felix Weinhardt. 2013. “Everybody Needs Good Neighbours? Evidence from Students’ Outcomes in England.” *Economic Journal*, 123: 831–874.

Gibbons, Stephen, Stephen Machin, and Olmo Silva. 2013. “Valuing school quality using boundary discontinuities.” *Journal of Urban Economics*, 75: 15–28.

Gould, Eric D., Victor Lavy, and M Daniele Paserman. 2004. “Immigrating to opportunity: Estimating the effect of school quality using a natural experiment on Ethiopians in Israel.” *The Quarterly Journal of Economics*, 119(2): 489–526.

Gould, Eric D., Victor Lavy, and M Daniele Paserman. 2011. “Sixty years after the magic carpet ride: The long-run effect of the early childhood environment on social and economic outcomes.” *The Review of Economic Studies*, 78(3): 938–973.

Goux, Dominique, and Eric Maurin. 2007. “Close neighbours matter: Neighbourhood effects on early performance at school.” *The Economic Journal*, 117(523): 1193–1215.

Hanushek, Eric A. 1986. “The economics of schooling: Production and efficiency in public schools.” *Journal of Economic Literature*, 24(3): 1141–1177.

Heckman, James J., John Eric Humphries, and Gregory Veramendi. 2017. “The Non-Market Benefits of Education and Ability.” National Bureau of Economic Research, Inc NBER Working Papers 23896.

Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. “The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior.” *Journal of Labor economics*, 24(3): 411–482.

Houtenville, Andrew J., and Karen Smith Conway. 2008. “Parental Effort, School Resources, and Student Achievement.” *Journal of Human Resources*, 43(2): 437–453.

Hoxby, Caroline. 2015. “Computing the value-added of american postsecondary institutions.” *Internal Revenue Service, US Department of the Treasury, Washington, DC.*

Imberman, Scott A., and Michael F. Lovenheim. 2016. “Does the market value value-added? Evidence from housing prices after a public release of school and teacher value-added.” *Journal of Urban Economics*, 91: 104–121.

Jackson, C. Kirabo. 2010. “Do Students Benefit From Attending Better Schools?: Evidence From Rule-based Student Assignments in Trinidad and Tobago.” *The Economic Journal*, 120(549).

Jackson, C. Kirabo. 2016. “What Do Test Scores Miss? The Importance Of Teacher Effects On Non-Test Score Outcomes.” National Bureau of Economic Research, Inc NBER Working Papers 22226.

Jacob, Brian A. 2004. “Public housing, housing vouchers, and student achievement: Evidence from public housing demolitions in Chicago.” *The American Economic Review*, 94(1): 233–258.

Jann, Ben. 2017. “KMATCH: Stata module for multivariate-distance and propensity-score matching.” *Statistical Software Components, Boston College Department of Economics.*

Kane, Thomas J., and Douglas O. Staiger. 2008. “Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation.” National Bureau of Economic Research 14607.

Kane, Thomas J., Stephanie K. Riegg, and Douglas O. Staiger. 2006. “School quality, neighborhoods, and housing prices.” *American Law and Economics Review*, 8(2): 183–212.

Katz, Lawrence F. 2015. “Reducing Inequality: Neighborhood and School Interventions.” *Focus*, 31: 12–17.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. “Experimental analysis of neighborhood effects.” *Econometrica*, 75(1): 83–119.

Lapierre, David, Pierre Lefebvre, and Philip Merrigan. 2016. “Long term educational attainment of private high school students in Québec: Estimates of treatment effects from longitudinal data.” Research Group on Human Capital Working Papers Series 16-02.

Lavecchia, Adam M., Heidi Liu, and Philip Oreopoulos. 2014. “Behavioral economics of education: Progress and possibilities.” National Bureau of Economic Research.

Lee, David S., and Thomas Lemieux. 2010. “Regression discontinuity designs in economics.” *Journal of Economic Literature*, 48(2): 281–355.

Lefebvre, Pierre, Philip Merrigan, and Matthieu Verstraete. 2011. “Public subsidies to private schools do make a difference for achievement in mathematics: Longitudinal evidence from Canada.” *Economics of Education Review*, 30(1): 79–98.

Ludwig, Jens, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, and Lisa Sanbonmatsu. 2013. “Long-Term Neighborhood Effects on Low-Income Families: Evidence from Moving to Opportunity." *The American Economic Review*, 103(3): 226–231.

McCrary, Justin. 2008. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of econometrics*, 142(2): 698–714.

Milligan, Kevin, Enrico Moretti, and Philip Oreopoulos. 2004. "Does education improve citizenship? Evidence from the United States and the United Kingdom." *Journal of Public Economics*, 88(9-10): 1667–1695.

Molitor, David. 2018. "The Evolution of Physician Practice Styles: Evidence from Cardiologist Migration." *American Economic Journal: Economic Policy*, 10(1): 326–356.

Moretti, Enrico. 2004. "Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data." *Journal of econometrics*, 121(1-2): 175–212.

Oreopoulos, Philip. 2003. "The Long-Run Consequences Of Living In A Poor Neighborhood." *The Quarterly Journal of Economics*, 118(4): 1533–1575.

Oreopoulos, Philip. 2008. "Neighborhood Effects in Canada: A Critique." *Canadian Public Policy-Analyse de politiques*, 34(2).

Oreopoulos, Philip. 2012. "Moving Neighborhoods Versus Reforming Schools: A Canadian's Perspective." *Cityscape*, 207–212.

Oreopoulos, Philip, and Kjell G. Salvanes. 2011. "Priceless: The Nonpecuniary Benefits of Schooling." *Journal of Economic Perspectives*, 25(1): 159–184.

Oreopoulos, Philip, and Uros Petronijevic. 2013. "Making College Worth It: A Review of Research on the Returns to Higher Education." National Bureau of Economic Research, Inc NBER Working Papers 19053.

Pei, Zhuan, Jorn-Steffen Pischke, and Hannes Schwandt. 2017. "Poorly Measured Confounders Are More Useful On The Left Than On The Right." National Bureau of Economic Research, Inc NBER Working Papers 23232.

Pop-Eleches, Cristian, and Miguel Urquiola. 2013. "Going to a Better School: Effects and Behavioral Responses." *American Economic Review*, 103(4): 1289–1324.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, schools, and academic achievement." *Econometrica*, 73(2): 417–458.

Rothstein, Jesse. 2006. "Good principals or good peers? Parental valuation of school characteristics, Tiebout equilibrium, and the incentive effects of competition among jurisdictions." *The American Economic Review*, 96(4): 1333–1350.

Rothstein, Jesse. 2017. "Inequality of Educational Opportunity? Schools as Mediators of the Intergenerational Transmission of Income."

Sharkey, Patrick, and Jacob W. Faber. 2014. "Where, when, why, and for whom do residential contexts matter? Moving away from the dichotomous understanding of neighborhood effects." *Annual Review of Sociology*, 40: 559–579.

Sykes, Brooke, and Sako Musterd. 2011. "Examining neighbourhood and school effects simultaneously: what does the Dutch evidence show?" *Urban Studies*, 48(7): 1307–1331.

Todd, Petra E., and Kenneth I. Wolpin. 2003. "On The Specification and Estimation of The Production Function for Cognitive Achievement." *Economic Journal*, 113(485): 3–33.

Weinhardt, Felix. 2014. "Social housing, neighborhood quality and student performance." *Journal of Urban Economics*, 82: 12–31.

Wodtke, Geoffrey T., and Matthew Parbst. 2017. "Neighborhoods, Schools, and Academic Achievement: A Formal Mediation Analysis of Contextual Effects on Reading and Mathematics Abilities." *Demography*.

Zimmerman, Seth D. 2014. "The returns to college admission for academically marginal students." *Journal of Labor Economics*, 32(4): 711–754.

For Online Publication

A Appendix Tables and Figures

Figure A1: Quebec's Education System

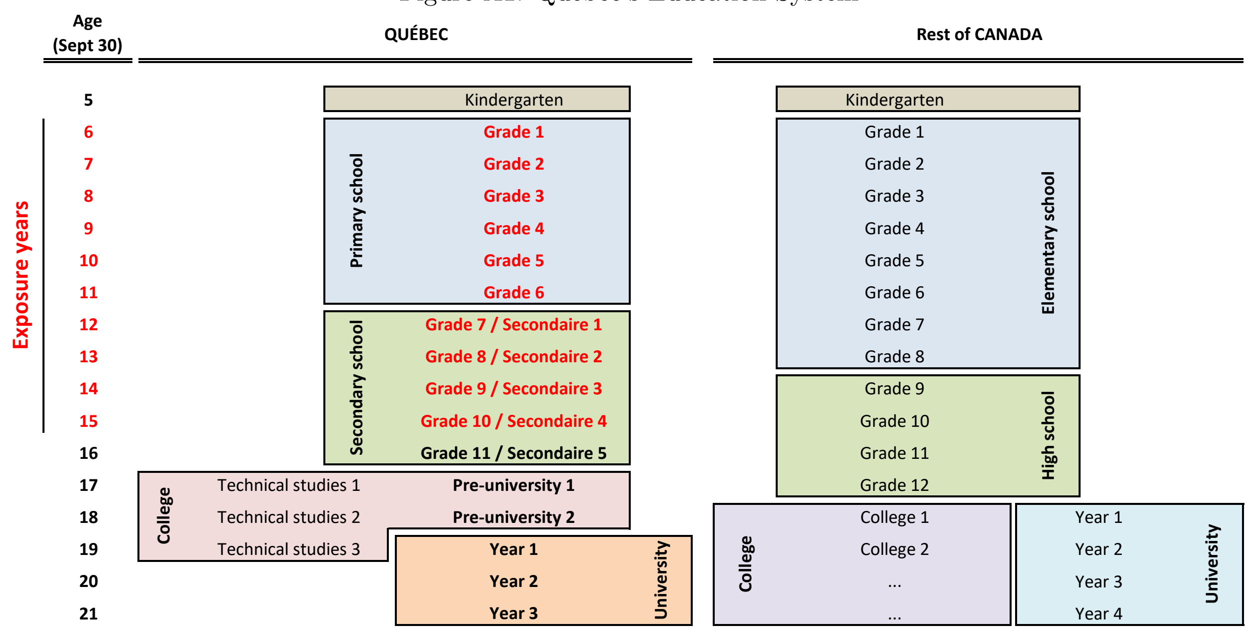
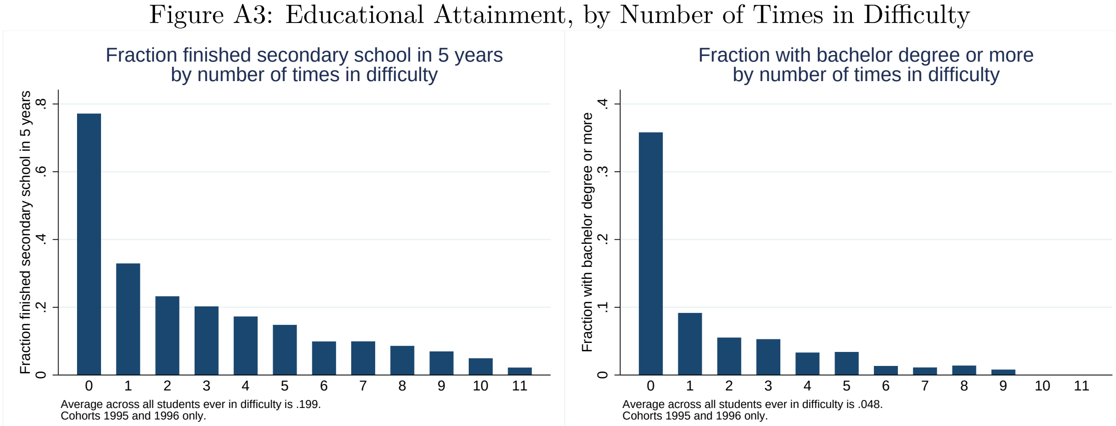


Figure A2: Catchment Areas and Census Tracts in Eastern Montreal (2001)

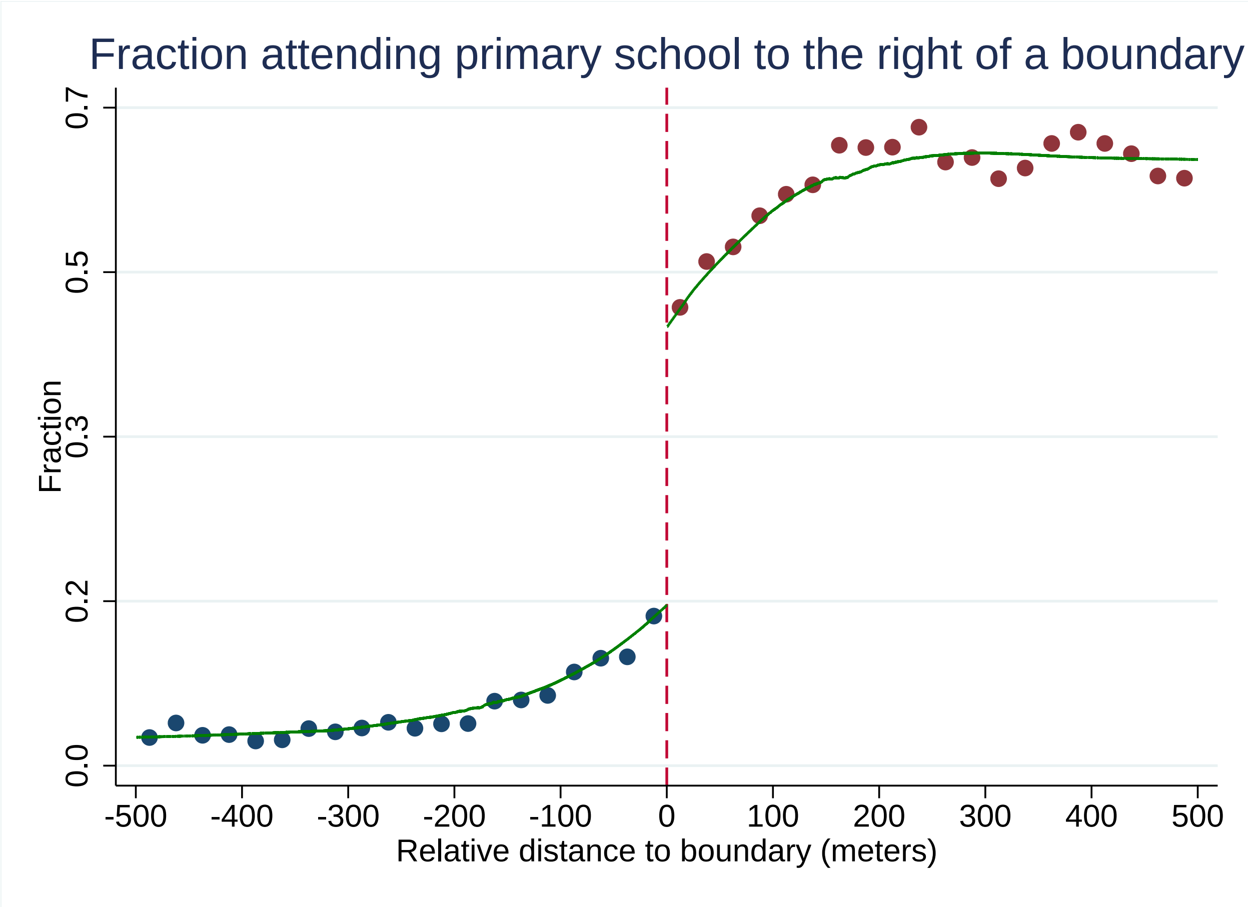
Notes: Colored areas indicate French primary school catchment areas as of 2001. Red lines denote census tracts.

Figure A3: Educational Attainment, by Number of Times in Difficulty



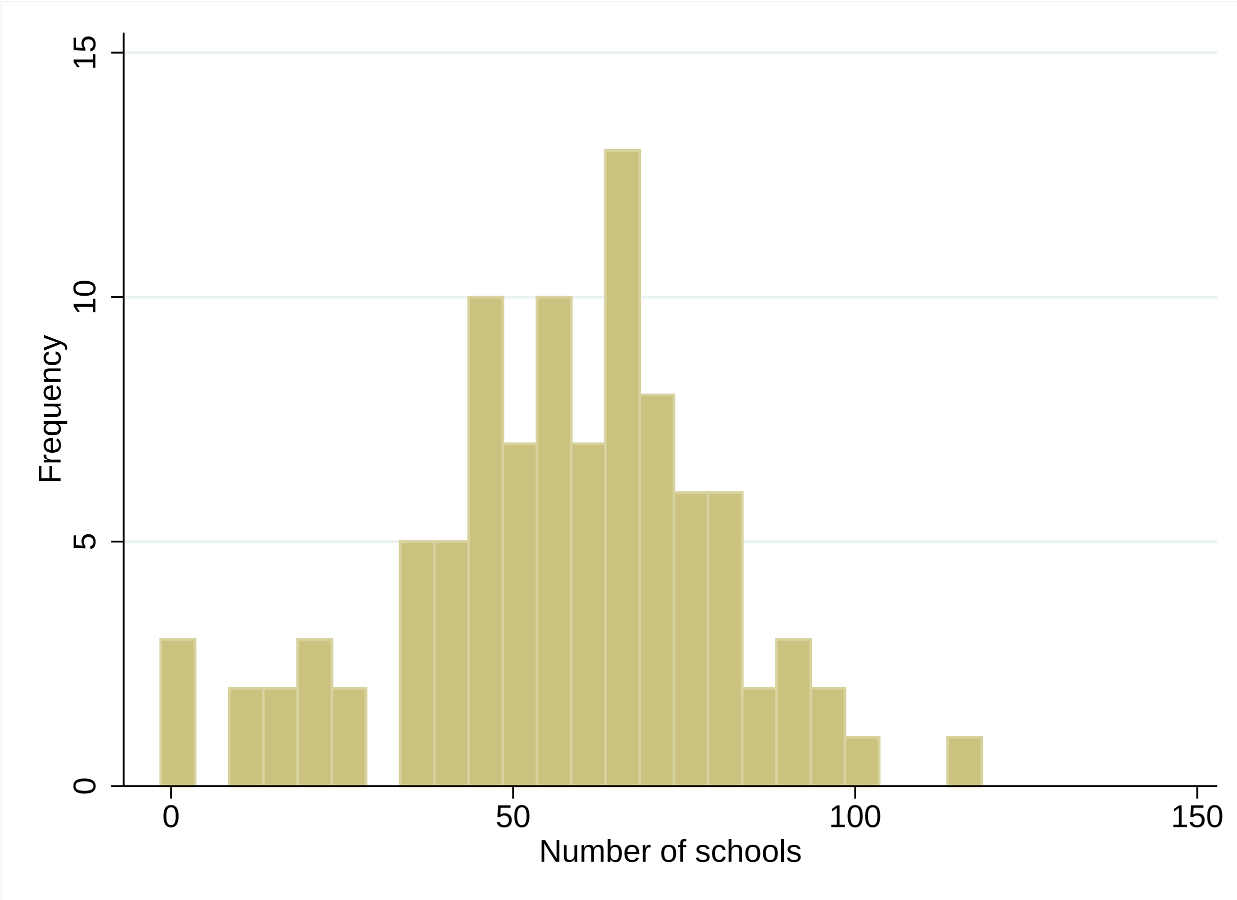
Notes: Sample restricted to students from the 1995 and 1996 cohorts. Too few students of the later cohorts have completed a bachelor degree by 2014-2015 to analyze this outcome for these students.

Figure A4: Discontinuity in School Attendance At Boundaries



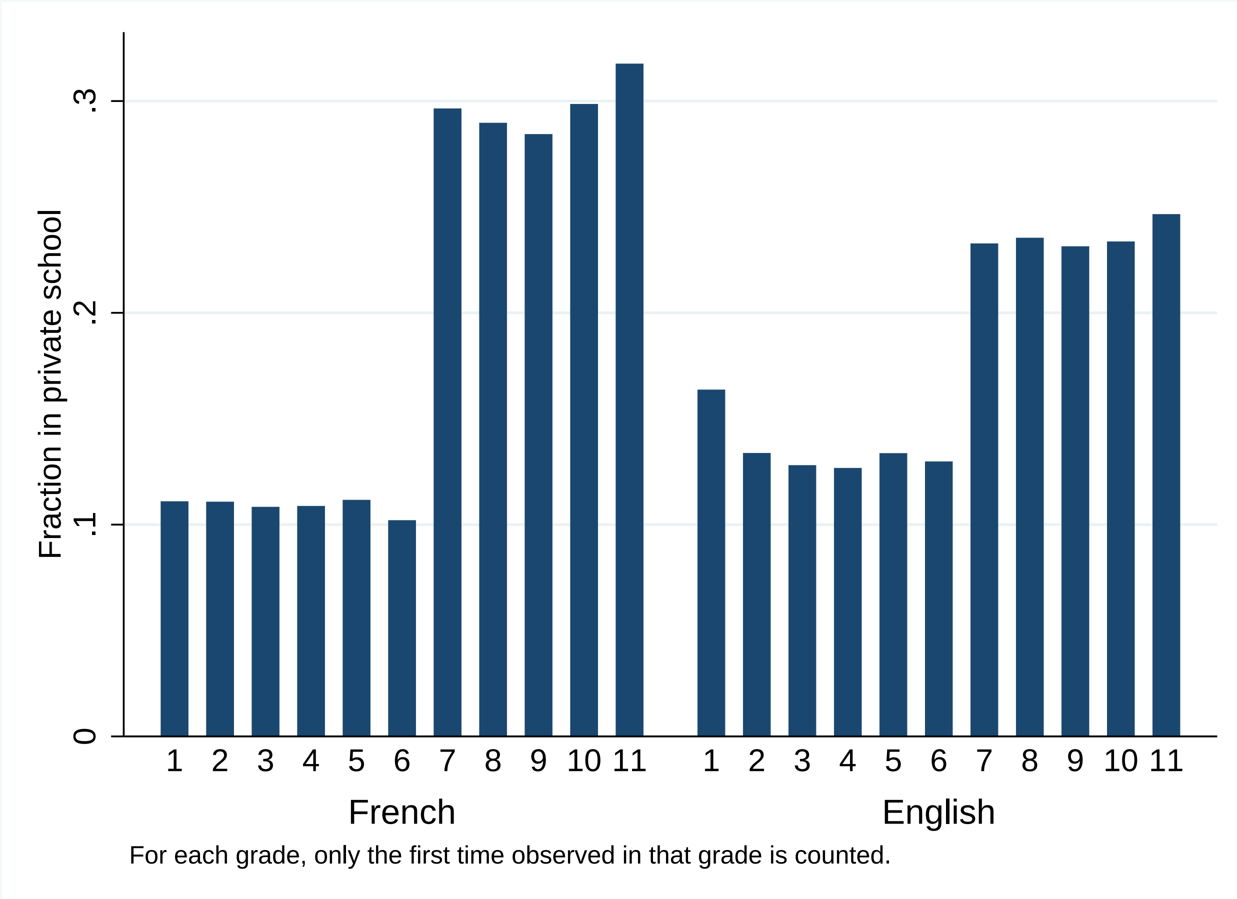
Notes: For each French primary school boundary, one neighborhood school is randomly assigned to the right. The figure shows the fraction of students enrolled in that school, by distance to the boundary. Students at positive distance are assigned the random chose default school. Students at negative distances are assigned to a school other than the one to the right. Attendance recorded at baseline (grade 1). Sample is restricted to students in French schools.

Figure A5: Distribution of School Choice Across FSAs



Notes: The histogram shows the distribution of FSAs by number of different primary schools attend by its residents. School attendance measured at baseline (i.e. first enrollment in grade 1).

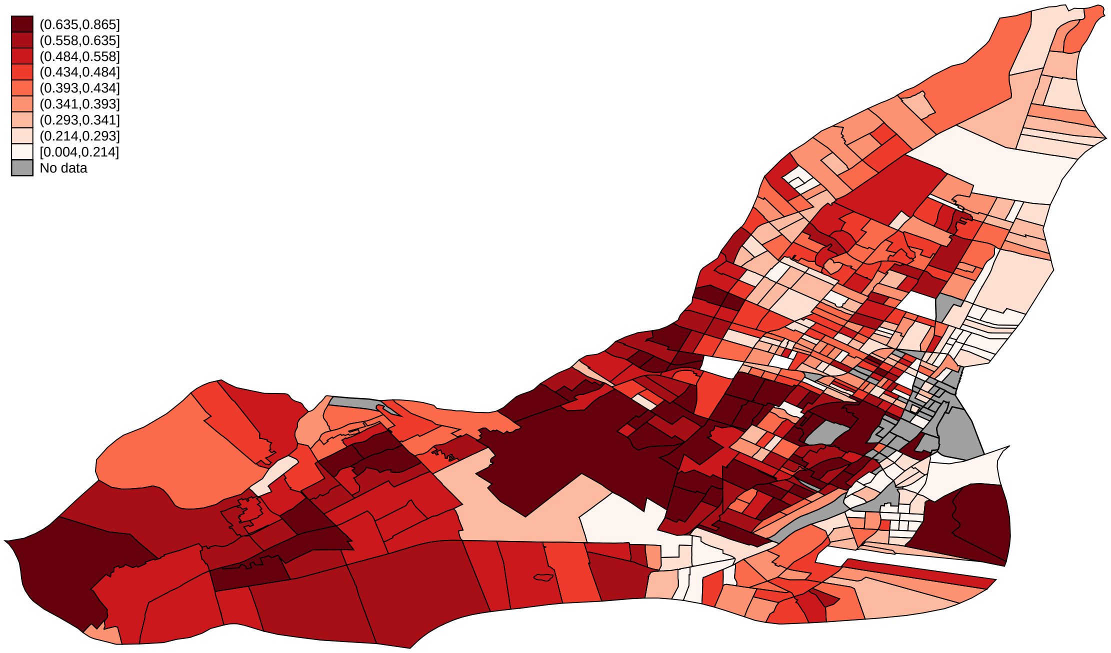
Figure A6: Fraction in Private Schools, by Grade and Language of Instruction



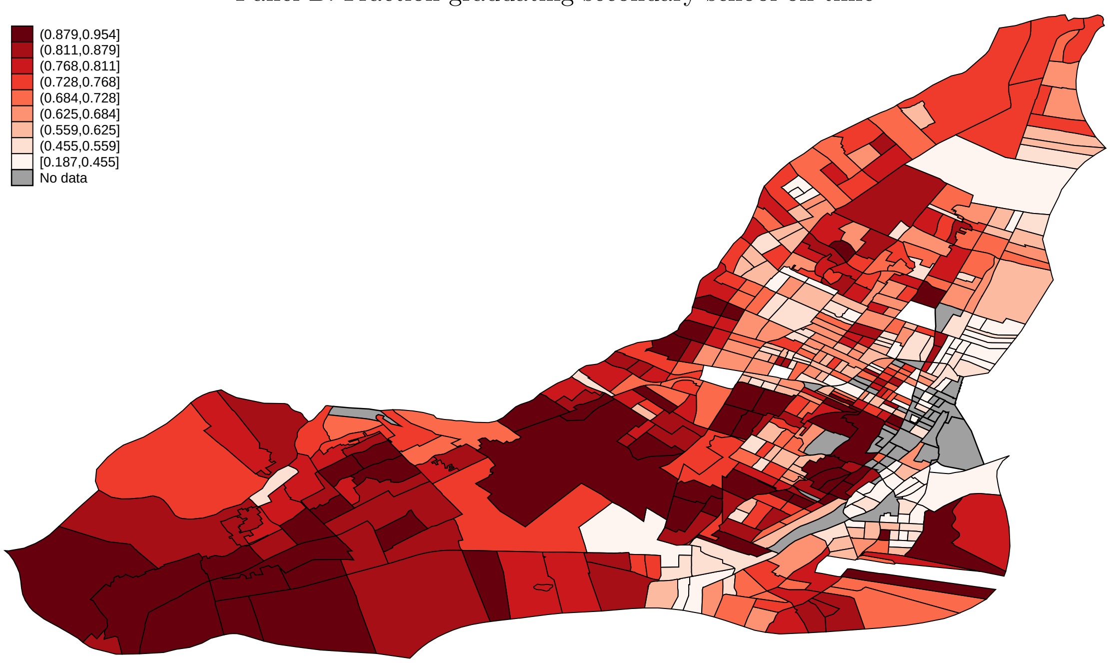
Notes: Statistics calculated over main analytical sample of 92,764 students. Data shown separately for students in French and English schools.

Figure A7: Spatial Variation in Educational Outcomes - Census Tract Level

Panel A: Fraction ever enrolled in university

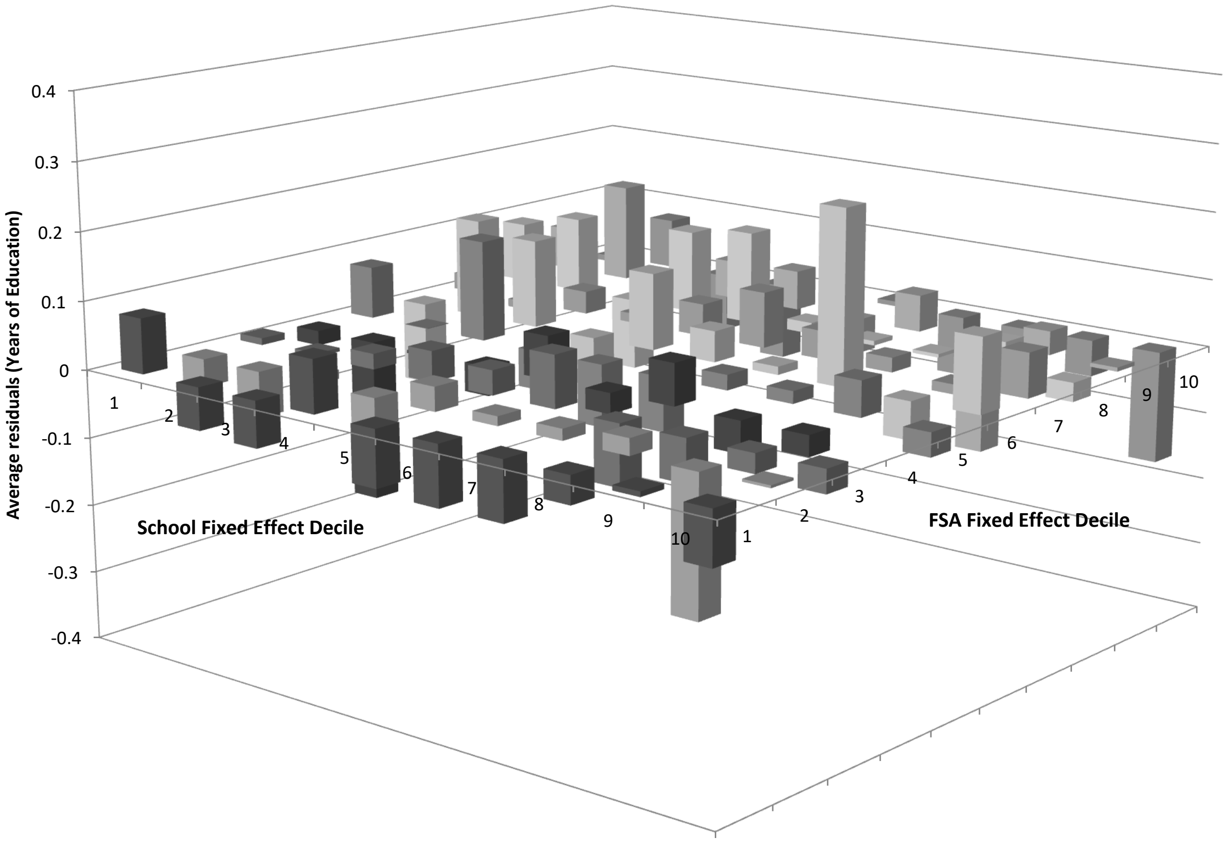


Panel B: Fraction graduating secondary school on time



Notes: Statistics based on permanent residents (students who always resided in the same census tract). Outcomes are adjusted for cohort effects. Data for census tracts with fewer than 10 permanent residents are not shown (no data).

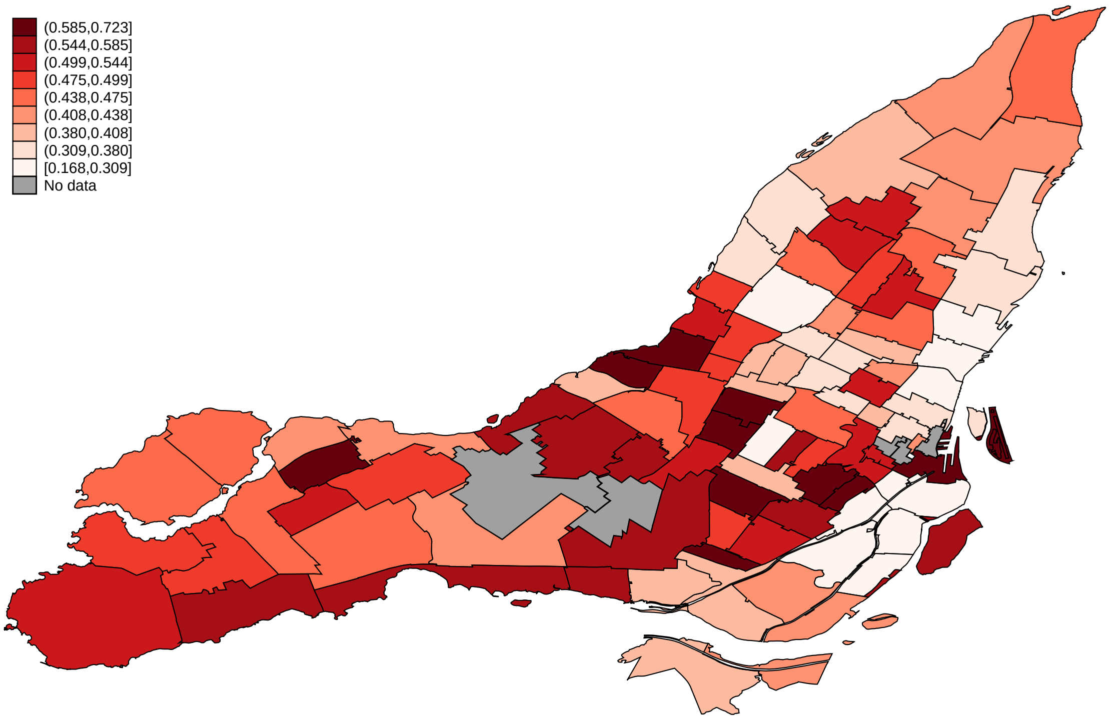
Figure A8: Mean Years of Education (Residuals), by School and FSA Deciles



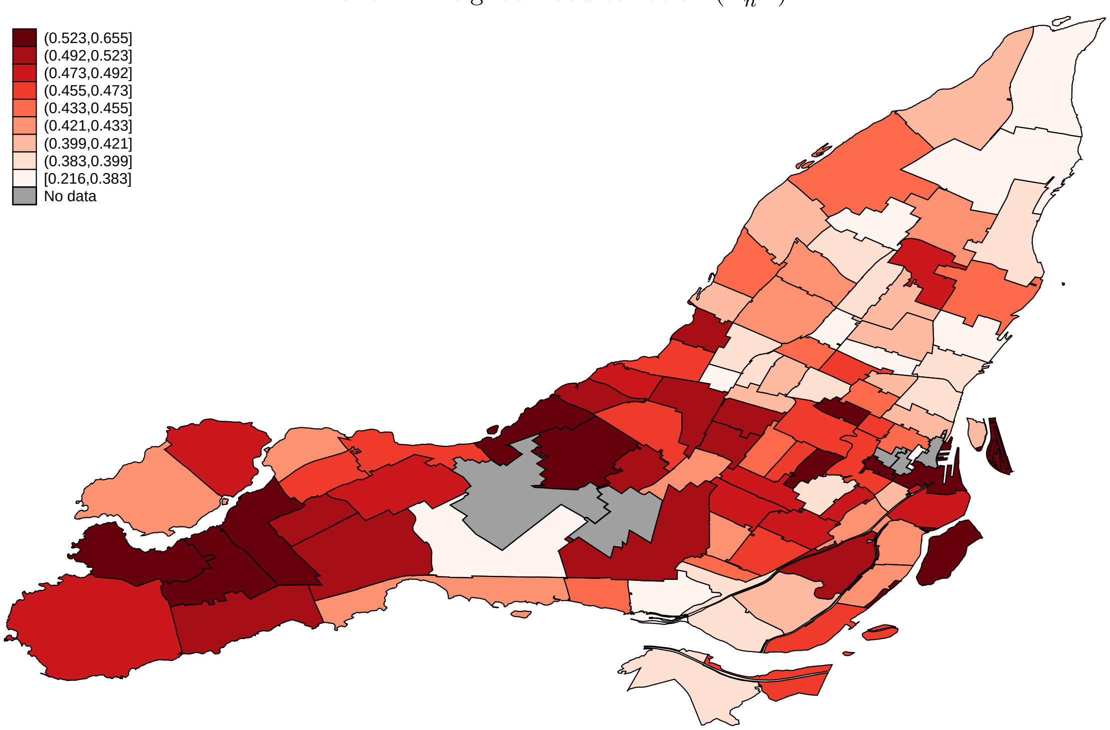
Notes: Residuals extracted from the estimation of a two-way fixed effect model, and correspond to the estimates reported in column (6) of Table 1. The figure is constructed by slicing the distributions of school and FSA fixed effects into deciles, and then calculating the average residuals in each school-by-neighborhood decile cell.

Figure A9: Spatial Variation in $\bar{\Omega}_n^{PR}$ and $\bar{\Lambda}_n^{PR}$, for University Enrollment

Panel A: School variation ($\bar{\Omega}_n^{PR}$)

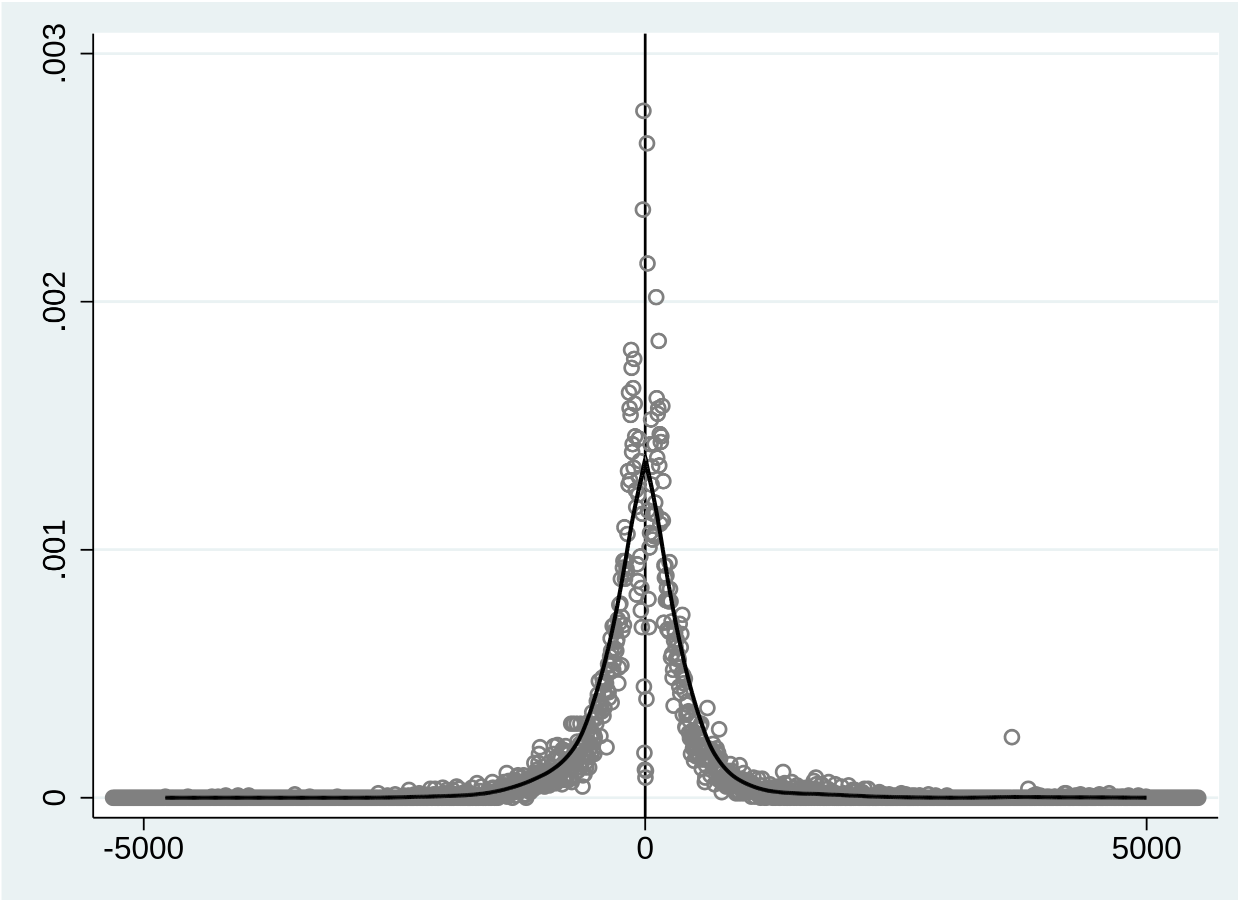


Panel B: Neighborhood variation ($\bar{\Lambda}_n^{PR}$)



Notes: Statistics based on permanent residents. Outcomes are adjusted for cohort effects. To ease the interpretation, the student-level fixed effects used to compute $\bar{\Omega}_n^{PR}$ and $\bar{\Lambda}_n^{PR}$ were first re-centered around the unconditional university enrollment rate for the the full sample. Data for FSAs with fewer than 10 permanent residents are not shown.

Figure A10: Density Plot around French Primary School Boundaries

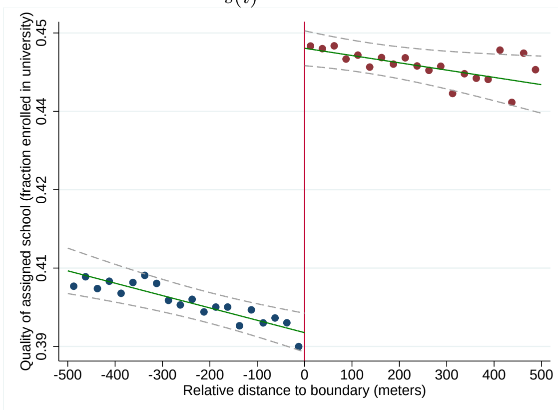


Notes: Figure produced with the Stata package DCdensity.ado, which implements the test derived in McCrary (2008). The x-axis shows distance relative to the nearest boundary, in meters.

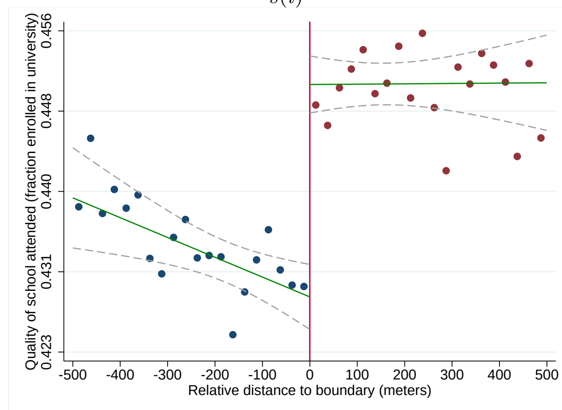
Figure A11: Discontinuities in School Quality at French Primary School Boundaries

All permanent residents

Panel A: Quality $\delta^P_{s(i)}$ of assigned French school

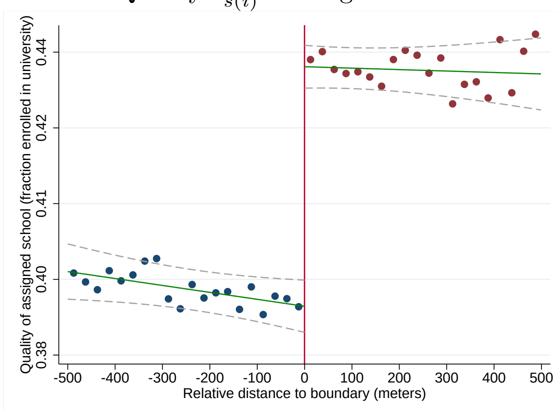


Panel B: Quality $\delta^P_{s(i)}$ of school attended

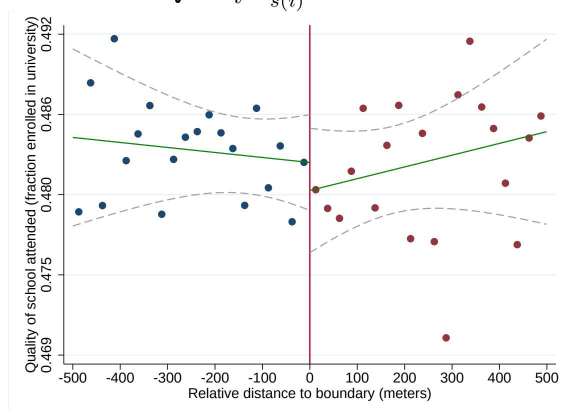


Students in English schools only (Placebo)

Panel C: Quality $\delta^P_{s(i)}$ of assigned French school



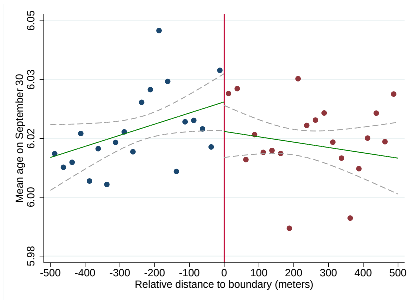
Panel D: Quality $\delta^P_{s(i)}$ of school attended



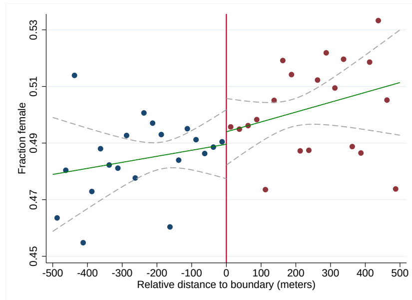
Notes: For each French primary school boundary, the neighborhood school with greater school quality – in terms of university enrollment – is assigned to the right. The variable on the vertical axis is first residualized on cohort, FSA, and boundary fixed effects. The figure shows the average school quality of schools attended by students at baseline, by distance to the boundary. Attendance recorded at baseline (grade 1). In Panels A and B, the sample includes all permanent residents, and in Panels C and D it is constituted of students enrolled in English schools only. For visual clarity, students living further than 500 meters away from their nearest boundary are excluded.

Figure A12: Balance of Covariates at Boundaries - School Quality in Terms of University Enrollment

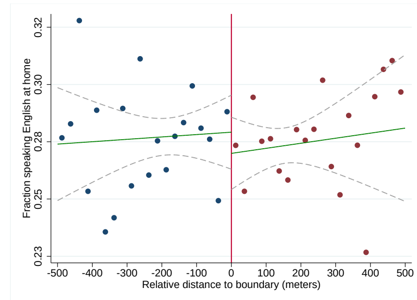
(a) Age



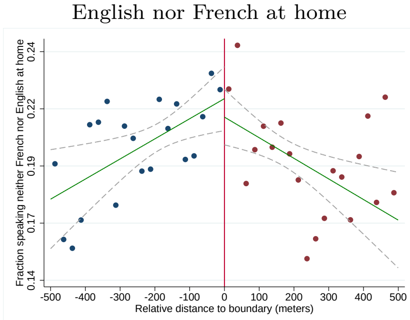
(b) Gender



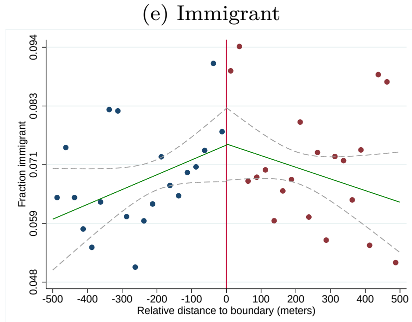
(c) Speaks English at home



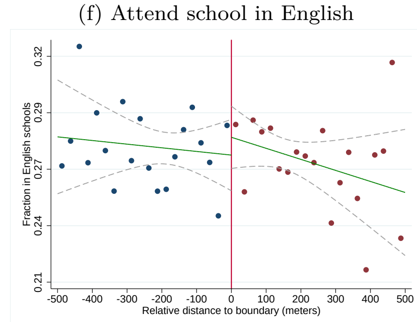
(d) Speaks neither English nor French at home



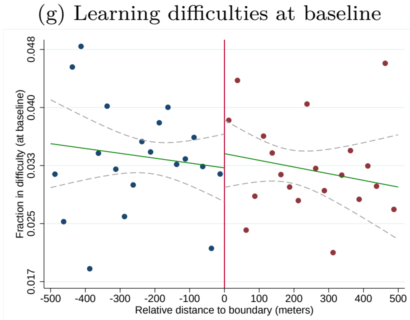
(e) Immigrant



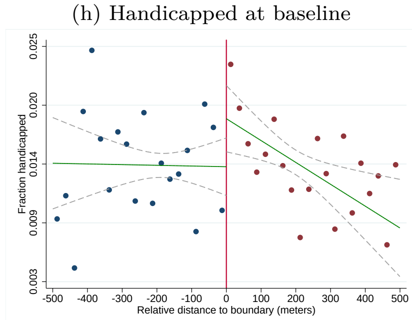
(f) Attend school in English



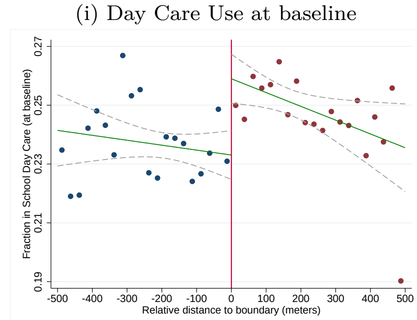
(g) Learning difficulties at baseline



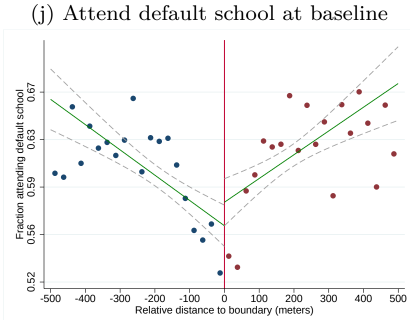
(h) Handicapped at baseline



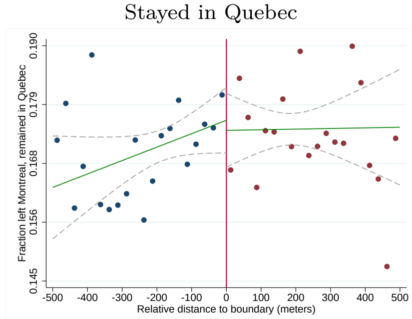
(i) Day Care Use at baseline



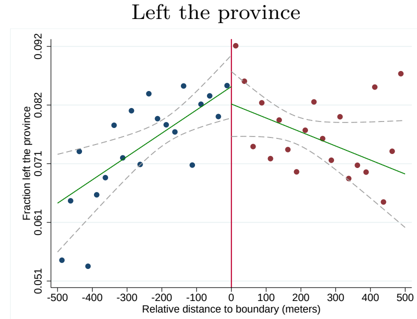
(j) Attend default school at baseline



(k) Attrition: Left Montreal, Stayed in Quebec



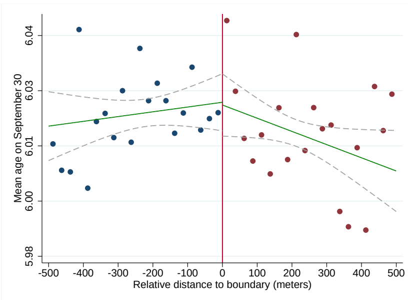
(l) Attrition: Left the province



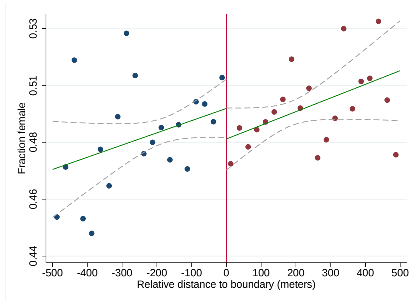
Notes: In panels (a) to (j), the sample is restricted to permanent residents. In panels (k) and (l), there is no sample restriction, hence all students in the database are included. For each boundary, students assigned the default school with the highest fixed effect $\delta^P_{s(i)}$ (measured in units of university enrollment) are at positive distances. Variables are first residualized on cohort, boundary and FSA fixed effects. Standard errors are clustered at the boundary level. For visual clarity, students living further than 500 meters away from their nearest boundary are excluded.

Figure A13: Balance of Covariates at Boundaries - School Quality in Terms of *DES* in 5 Years

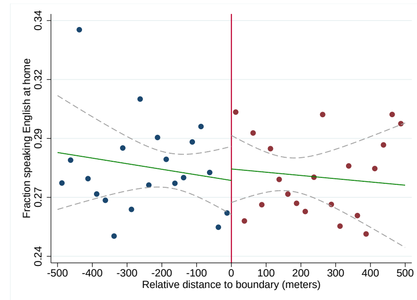
(a) Age



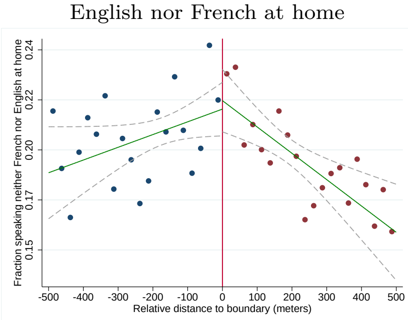
(b) Gender



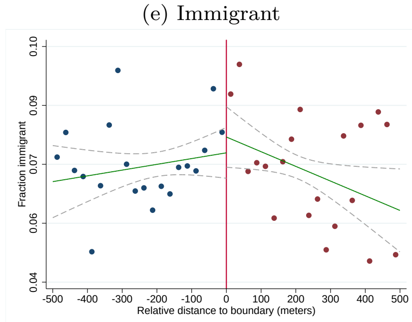
(c) Speaks English at home



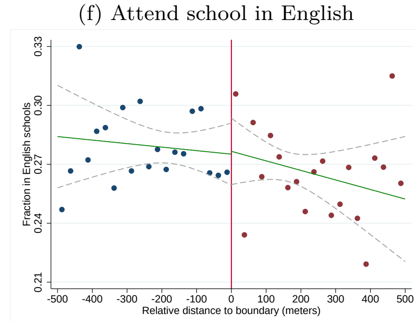
(d) Speaks neither English nor French at home



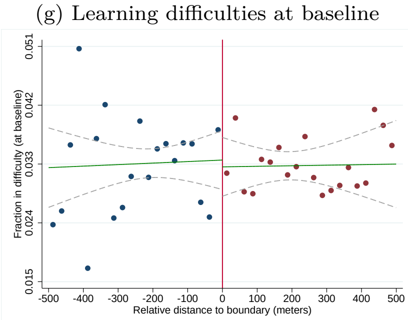
(e) Immigrant



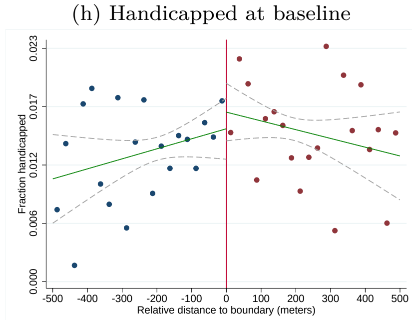
(f) Attend school in English



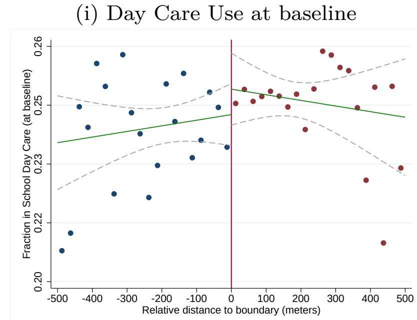
(g) Learning difficulties at baseline



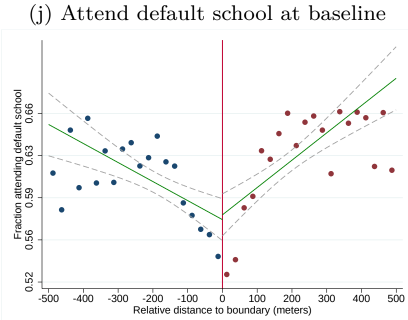
(h) Handicapped at baseline



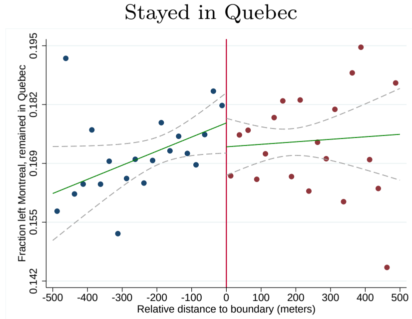
(i) Day Care Use at baseline



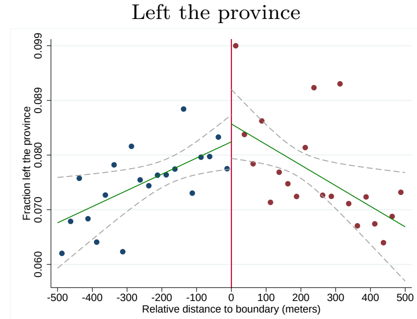
(j) Attend default school at baseline



(k) Attrition: Left Montreal, Stayed in Quebec



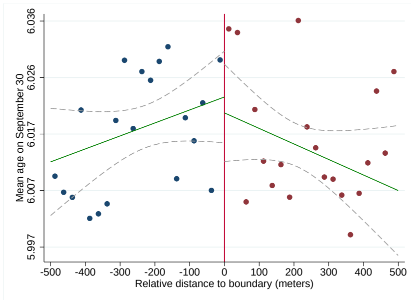
(l) Attrition: Left the province



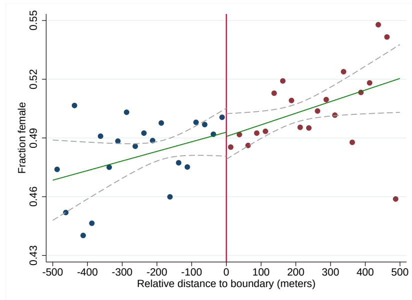
Notes: In panels (a) to (j), the sample is restricted to permanent residents. In panels (k) and (l), there is no sample restriction, hence all students in the database are included. For each boundary, students assigned the default school with the highest fixed effect $\delta^P_{s(i)}$ (measured in units of timely secondary school graduation) are at positive distances. Variables are first residualized on cohort, boundary and FSA fixed effects. Standard errors are clustered at the boundary level. For visual clarity, students living further than 500 meters away from their nearest boundary are excluded.

Figure A14: Balance of Covariates at Boundaries - School Quality in Terms of Years of Education

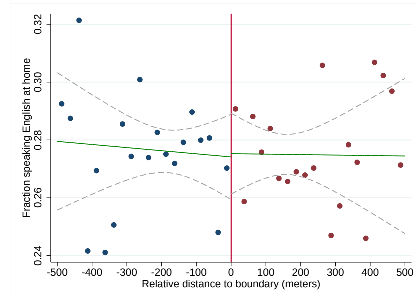
(a) Age



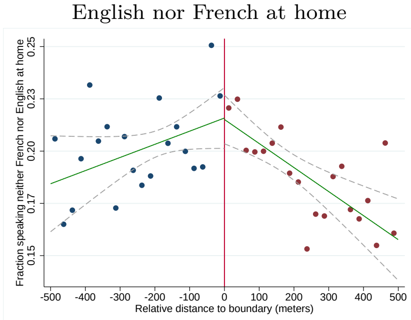
(b) Gender



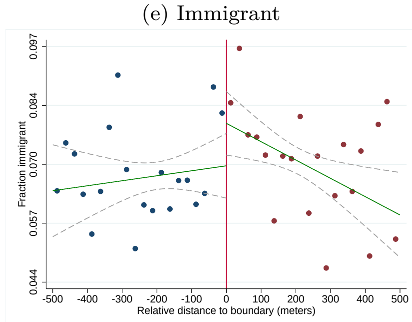
(c) Speaks English at home



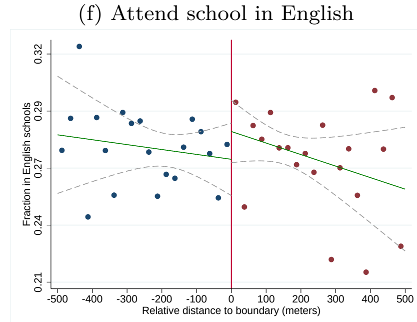
(d) Speaks neither English nor French at home



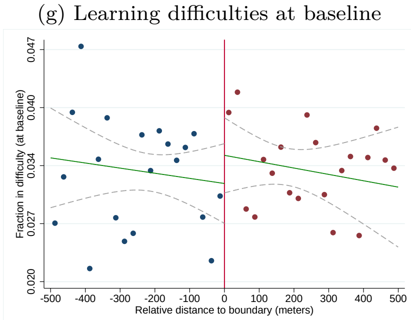
(e) Immigrant



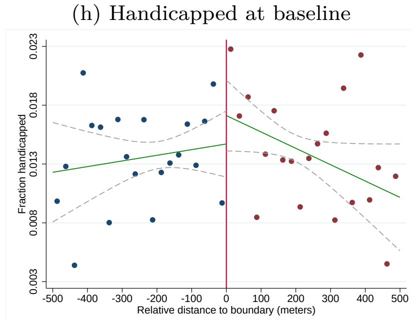
(f) Attend school in English



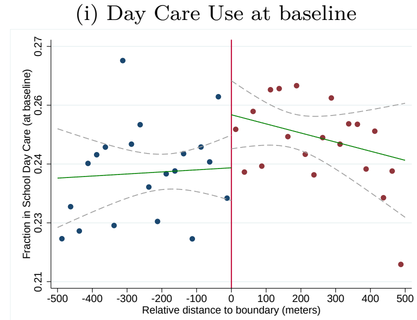
(g) Learning difficulties at baseline



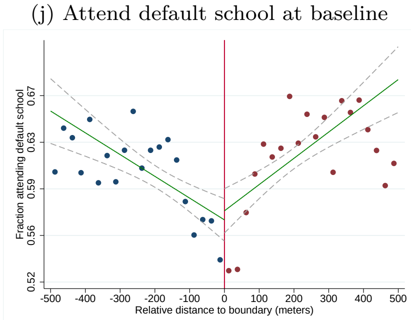
(h) Handicapped at baseline



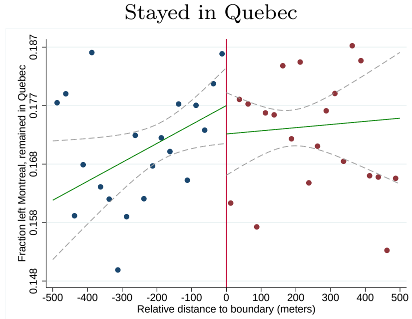
(i) Day Care Use at baseline



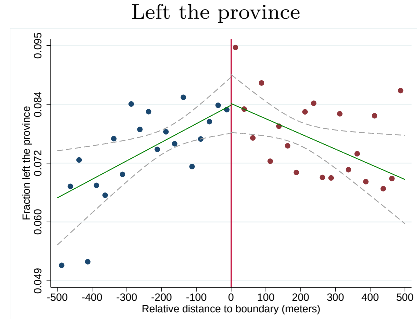
(j) Attend default school at baseline



(k) Attrition: Left Montreal, Stayed in Quebec

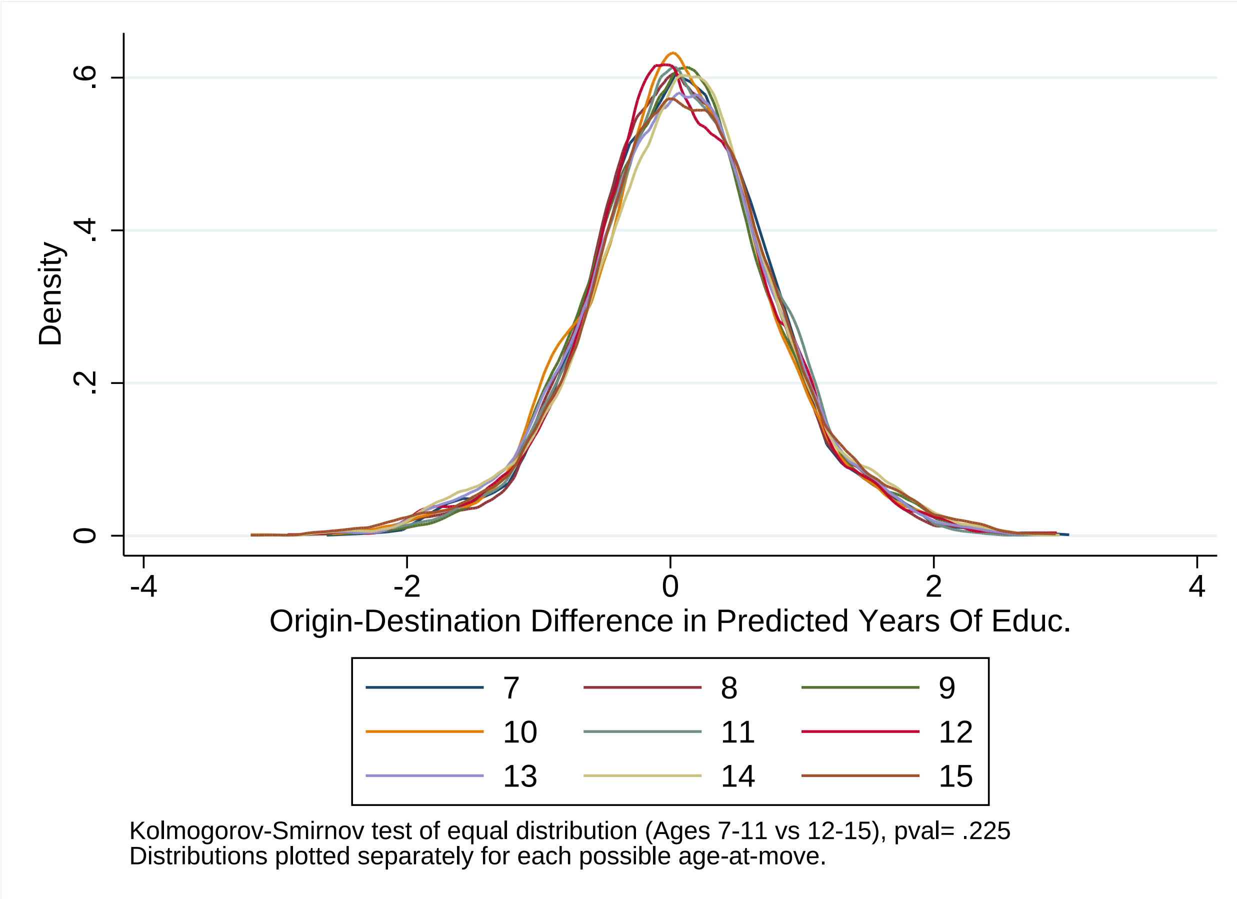


(l) Attrition: Left the province



Notes: In panels (a) to (j), the sample is restricted to permanent residents. In panels (k) and (l), there is no sample restriction, hence all students in the database are included. For each boundary, students assigned the default school with the highest fixed effect $\delta^P_{s(i)}$ (measured in units of university enrollment) are at positive distances. Variables are first residualized on cohort, boundary and FSA fixed effects. Standard errors are clustered at the boundary level. For visual clarity, students living further than 500 meters away from their nearest boundary are excluded.

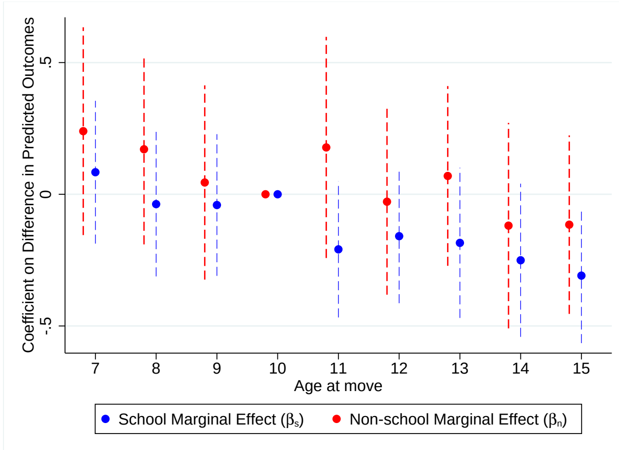
Figure A15: Distribution of $\Delta \bar{y}_{od}$ by Age-at-move



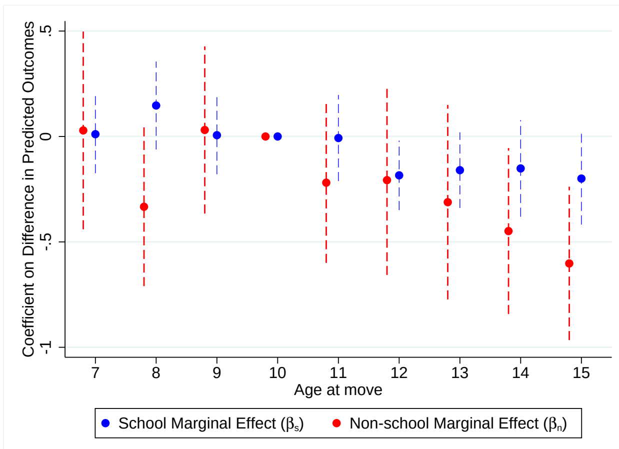
Notes: The kernel density of the distribution of $\Delta \bar{y}_{od}$ (in years of education) is plotted separated for each possible value of age-at-move.

Figure A16: Semi-parametric Partial Exposure Effects

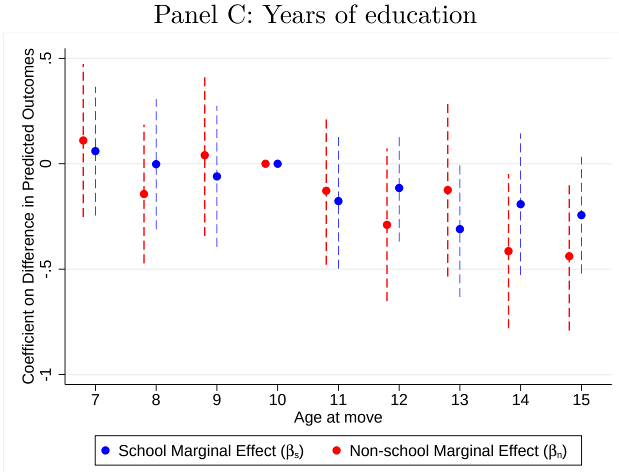
Panel A: University enrollment



Panel B: *DES* in 5 Years

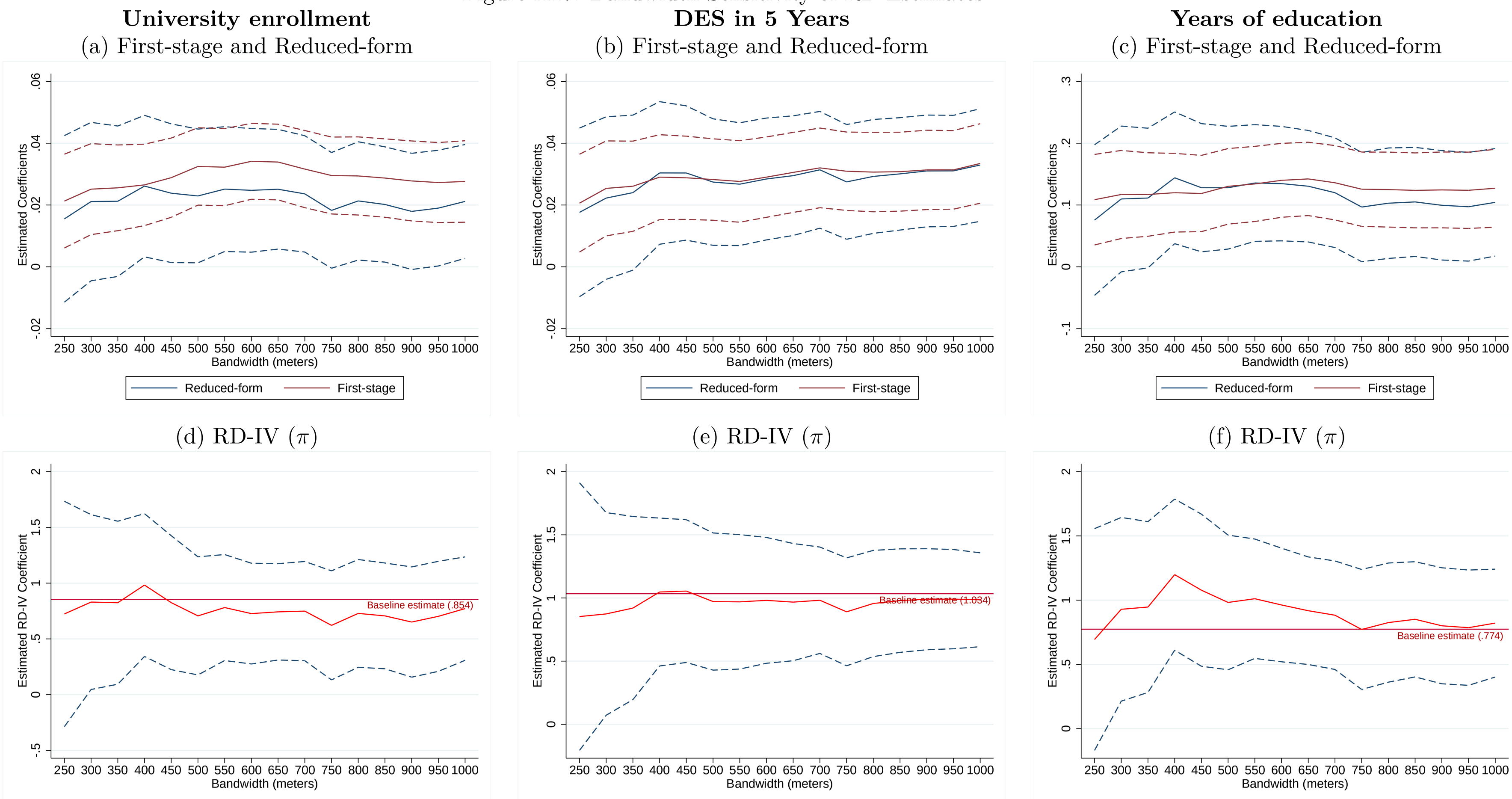


Panel C: Years of education



Notes: Notes: Sample includes all movers who remained within Montreal. Observation in FSAs with less than 10 permanent residents are omitted. Blue dots correspond to age-specific partial regression coefficients of y_i on $\pi\Delta\Omega_{od}$, and red dots are similarly defined for coefficients on $\Delta\bar{y}_{od}^{-s}$ in the following regression: $y_{icmod} = \sum_{m=7}^{15} \beta_{s,m} \left(\pi\Delta\Omega_{od} \times 1\{m_i = m\}\right) + \sum_{m=7}^{15} \beta_{n,m} \left(\Delta\bar{y}_{od}^{-s} \times 1\{m_i = m\}\right) + \gamma X_{icmod} + \alpha_{od} + \alpha_m + \alpha_c + \varepsilon_{icmod}$. Standard errors are clustered at the destination level.

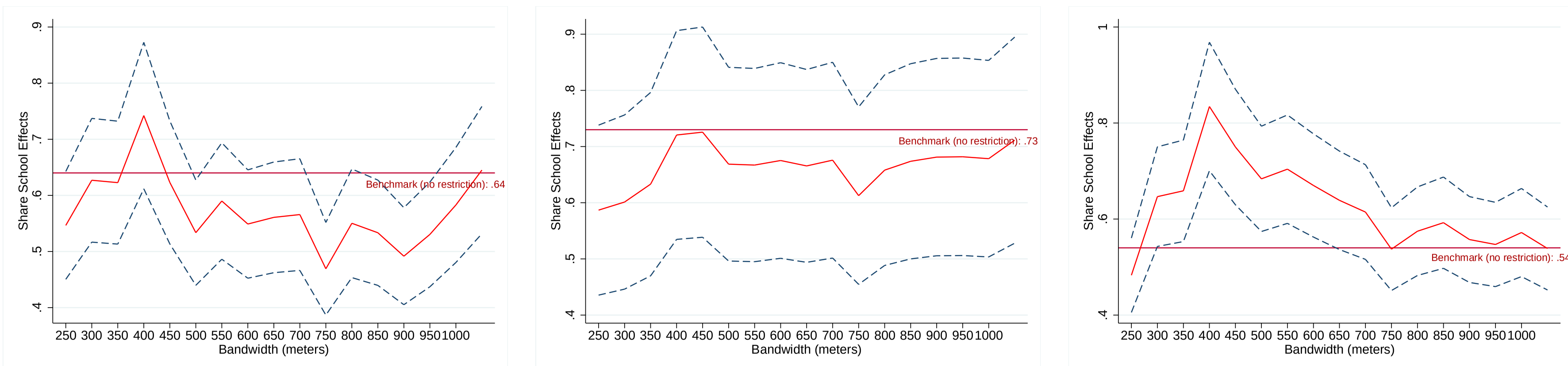
Figure A17: Bandwidth Sensitivity of RD Estimates



Notes: Panels (a) to (c) show first-stage and reduced-form RD coefficients for different bandwidth values. The dashed lines represent 95% confidence intervals with standard errors clustered at the boundary-level. Panels (d) to (f) show the associated RD-IV coefficients. The horizontal line shows the baseline RD-IV estimate under no bandwidth restriction.

Figure A18: Bandwidth Sensitivity of School Share S^{school}

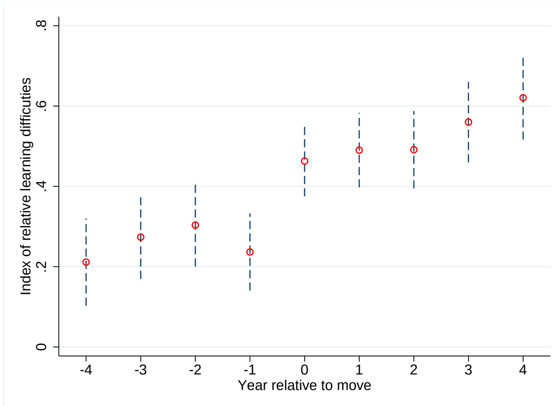
Panel A: University enrollment | Panel B: DES in 5 Years | Panel C: Years of education



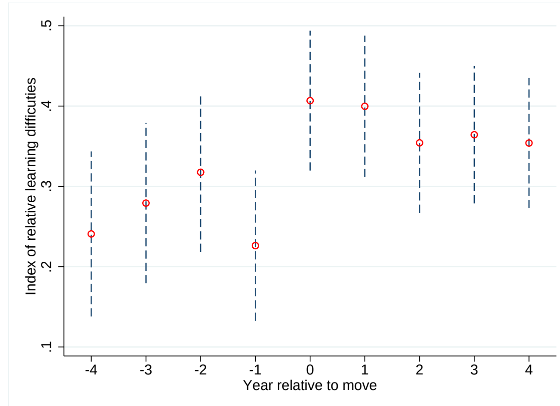
Notes: For each panel, the red line shows the share of total exposure effects due to school S^{school} for different RD bandwidth values. The dashed lines represent 95% confidence intervals with standard errors calculated by the delta method. The horizontal line shows the baseline estimate of the school share under no bandwidth restriction.

Figure A19: Index of Relative Learning Difficulties, by Years Relative to Move

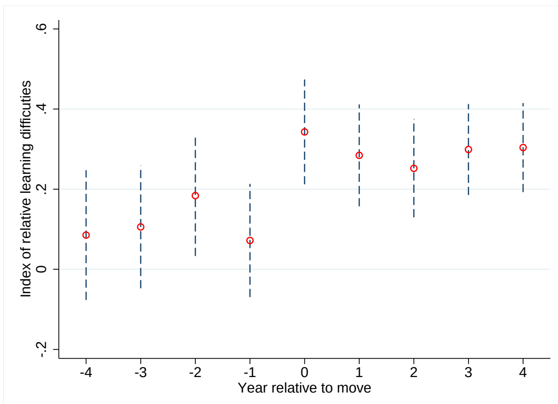
Panel A: No student fixed effects



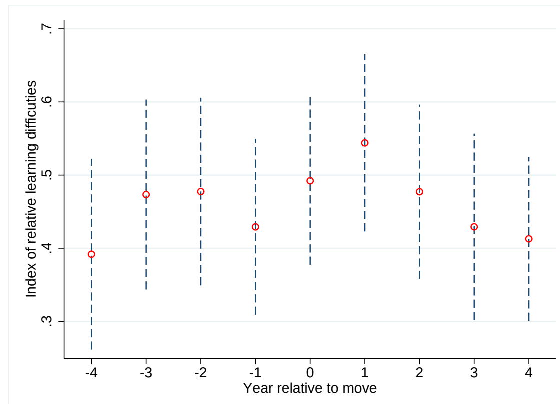
Panel B: With student fixed effects



Panel C: With student fixed effects
School switchers



Panel D: With student fixed effects
Non-switchers

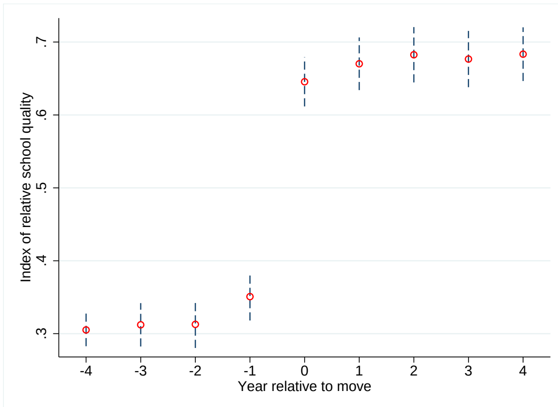


Notes: Standard errors are clustered at the individual level. The y-axis shows regression coefficients of $\sigma_{od(i,t)}$ on relative-time dummies, net of cohort, age and time (i.e. years since started grade 1) fixed effects. Observations outside the event window are included in the regression, so all coefficients are relative to omitted relative-time periods. Panel C includes only students who switched school the year they moved. Panel D includes movers who did not switch school the year they moved. Observations are weighted by $\left(\overline{Diff}_{dt} - \overline{Diff}_{ot}\right)^2$, and weights are normalized within time periods.

Figure A20: Index of Relative School Quality, by Years Relative to Move

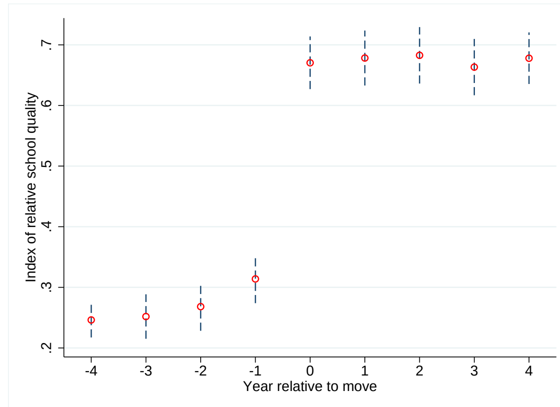
Panel A: University enrollment

All movers



Pre-post difference: 0.47 (s.e. 0.014)

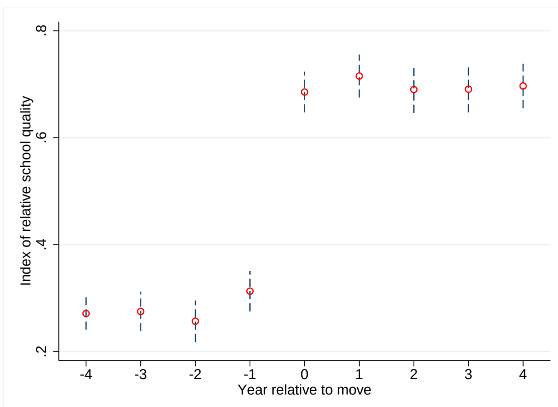
One-time movers



Pre-post difference: 0.54 (s.e. 0.018)

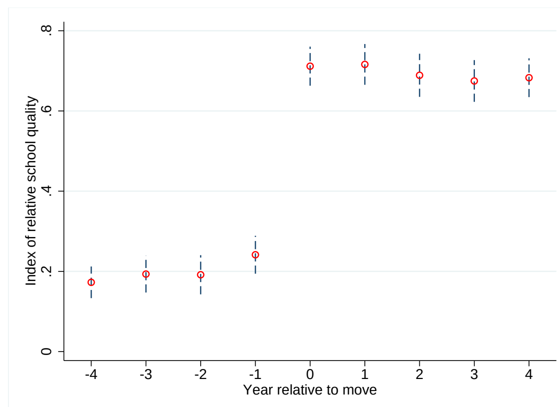
Panel B: DES in 5 Years

All movers



Pre-post difference: 0.56 (s.e. 0.019)

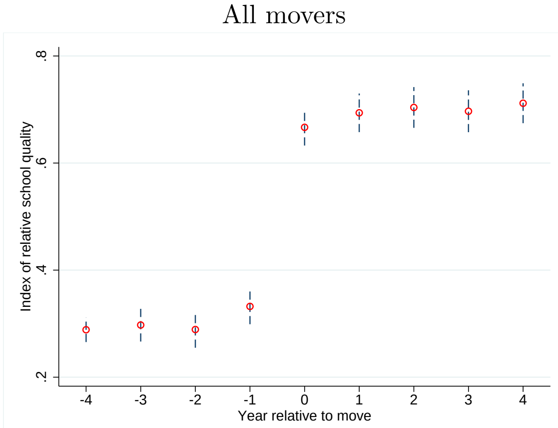
One-time movers



Pre-post difference: 0.65 (s.e. 0.024)

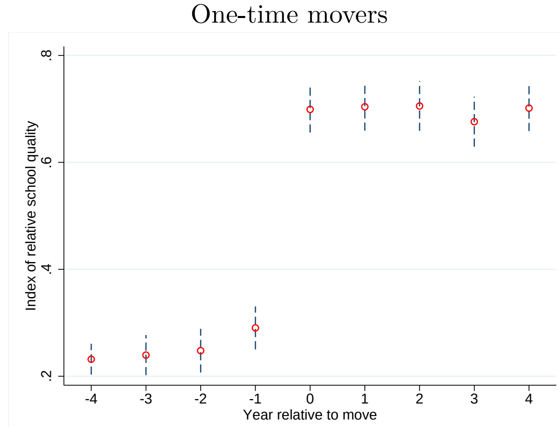
Panel C: Years of education

All movers



Pre-post difference: 0.52 (s.e. 0.015)

One-time movers



Pre-post difference: 0.59 (s.e. 0.019)

Notes: Standard errors are clustered at the individual level. The y-axis shows regression coefficients of $\sigma^{\psi}_{od(i,t)} = \frac{\delta_{s(i,t)} - \overline{\delta}_{s(o,t)}}{\overline{\delta}_{s(d,t)} - \overline{\delta}_{s(o,t)}}$ on relative-time dummies, net of cohort, age and time fixed effects. For each period t, $\overline{\delta}_{s(n,t)}$ is measured by the relevant average primary school fixed effects if student i was in primary school in that year. Secondary school fixed effects are used for remaining years. Observations outside the event window are included in the regression, so all coefficients shown on figures are relative to omitted relative-time periods. Pre-post differences are calculated over all years (no event window restriction). Observations are weighted by $\left(\overline{\delta}_{s(d,t)} - \overline{\delta}_{s(o,t)}\right)^2$, and weights are normalized within time periods.

Table A1: Descriptive Statistics

	All students	Permanent residents	Movers: Within Montreal	Movers: Left Montreal	Difference between (2) and (3)
	mean (s.d.)	mean (s.d.)	mean (s.d.)	mean (s.d.)	coef. (s.e.)
Variables	(1)	(2)	(3)	(4)	(5)
Female	0.49	0.49	0.49	0.49	0.001
	[0.500]	[0.500]	[0.500]	[0.500]	[0.004]
Age on September 30	6.02	6.01	6.04	6.02	-0.033
	[0.376]	[0.329]	[0.455]	[0.330]	[0.003]
Mother tongue: French	0.49	0.47	0.45	0.65	0.026
	[0.500]	[0.499]	[0.497]	[0.478]	[0.004]
Mother tongue: English	0.21	0.26	0.20	0.09	0.063
	[0.407]	[0.439]	[0.397]	[0.291]	[0.003]
Mother tongue: Other	0.30	0.27	0.36	0.26	-0.090
	[0.457]	[0.444]	[0.479]	[0.439]	[0.003]
Language at home: French	0.54	0.50	0.50	0.69	0.006
	[0.499]	[0.500]	[0.500]	[0.460]	[0.004]
Language at home: English	0.26	0.32	0.24	0.12	0.077
	[0.437]	[0.466]	[0.427]	[0.326]	[0.003]
Language at home: Other	0.21	0.18	0.26	0.18	-0.083
	[0.405]	[0.384]	[0.438]	[0.387]	[0.003]
Immigrant	0.10	0.07	0.14	0.11	-0.080
	[0.296]	[0.247]	[0.347]	[0.308]	[0.002]
Language at school: French	0.75	0.69	0.77	0.88	-0.073
	[0.433]	[0.461]	[0.423]	[0.324]	[0.003]
Uses School Day Care (baseline)	0.25	0.24	0.24	0.29	0.005
	[0.432]	[0.428]	[0.426]	[0.453]	[0.003]
In difficulty (baseline)	0.04	0.03	0.05	0.05	-0.016
	[0.193]	[0.170]	[0.209]	[0.219]	[0.001]
Handicapped (baseline)	0.01	0.01	0.02	0.01	-0.003
	[0.118]	[0.116]	[0.126]	[0.111]	[0.001]
Ever in difficulty by age 15	0.31	0.25	0.37	0.37	-0.116
	[0.462]	[0.431]	[0.482]	[0.483]	[0.003]
Students	92,764	44,912	31,526	16,326	76,438

Notes: The main sample excludes students who left Quebec's system before turning 16. Permanent residents are defined as students who always resided in the same FSA until the age of 15. Movers within Montreal are those who moved across FSAs at least once and were still living on the Island of Montreal at age 15. Movers who left Montreal were residing in the province of Quebec but outside the Island of Montreal at age 15.

Table A2: Summary Statistics: Educational Outcomes Across Cohorts

	All	Cohort 1995	1996	1998	2000	2001
Primary and secondary school outcomes						
Did not start secondary school on time	0.113	0.156	0.153	0.124	0.073	0.068
Secondary school diploma	0.760	0.755	0.752	0.759	0.767	0.765
Secondary school diploma in 5 years	0.610	0.600	0.587	0.609	0.630	0.621
No secondary school qualification	0.200	0.208	0.209	0.195	0.189	0.198
Post-secondary outcomes						
Ever enrolled in college	0.695	0.678	0.682	0.699	0.710	0.705
Enrolled in college by age 17	0.530	0.497	0.503	0.532	0.560	0.555
Ever enrolled in university	0.373	0.460	0.451	0.424	0.332	0.220
Enrolled in university by age 19	0.170	0.166	0.166	0.169	0.175	0.175
Bachelor degree or more	0.128	0.275	0.249	0.140	0.003	0.004
Educational attainment						
Number of years of education	12.810	13.247	13.200	13.066	12.517	12.119
Observations	92,764	16,969	18,067	18,777	19,125	19,826

Notes: The table shows cohort-specific average outcomes. These statistics exclude almost 1,000 individuals who enroll in a Quebec post-secondary institution at some point, but who had left the primary and secondary school system before turning 16 and therefore are excluded from the main sample.

Table A3: Variation Across FSAs and Schools - Empirical Bayes Estimates

	Outcome: DES in 5 years		University enrollment		Years of education	
	(1)	(2)	(3)	(4)	(5)	(6)
Student-level standard deviation of shrunk fixed effects:						
Schools	0.263	0.251	0.243	0.218	1.180	1.073
Neighborhoods (FSAs)	0.127	0.016	0.129	0.035	0.636	0.148
Dependent variable summary statistics:						
Mean	0.706		0.443		13.228	
Standard deviation	[0.456]		[0.497]		[2.113]	
Fixed effects estimated						
Separately	x		x		x	
Simultaneously		x		x		x
Number of students	44,912					
Number of primary schools	440					
Number of secondary schools	218					
Number of neighborhoods	95					

Notes: Sample restricted to students who always resided in the same FSA. School fixed effects are the sum of a primary and a secondary school fixed effect. To shrink estimates, I first calculate standard errors for each school and neighborhood fixed effect using bootstrap resampling (100 samples with replacement, clustering within primary school-secondary school-FSA cells). I then shrink estimates toward their means using the empirical Bayes procedure described in Chandra et al. (2016). Note that the reported empirical Bayes measures of school effects are shrunk estimates of the sum of primary and secondary school effects, not the sum of shrunk estimates of primary school and shrunk estimates of secondary school effects.

Table A4: Balance of Covariates at Boundaries

Outcome used to assign *HighSide:*	University Enrollment	DES in 5 years	Years of education
	(1)	(2)	(3)
Covariates			
Age	-0.0091	-0.0075	-0.0044
	(0.0057)	(0.0059)	(0.0058)
Gender	0.0204	0.0093	0.0176
	(0.0071)	(0.0072)	(0.0072)
Speaks English at Home	-0.0041	0.0042	-0.0019
	(0.0114)	(0.0106)	(0.0114)
Speaks neither French nor English at Home	-0.0095	-0.0078	-0.0153
	(0.0086)	(0.0085)	(0.0086)
Immigrant	0.0004	0.0012	0.0029
	(0.0046)	(0.0046)	(0.0044)
Attend school in English	-0.0020	-0.0099	-0.0004
	(0.0120)	(0.0115)	(0.0121)
Learning difficulties at baseline	0.0005	-0.0010	0.0022
	(0.0030)	(0.0031)	(0.0031)
Handicapped at baseline	0.0026	0.0026	0.0018
	(0.0017)	(0.0017)	(0.0017)
Day Care Useat baseline	0.0205	0.0065	0.0123
	(0.0067)	(0.0068)	(0.0068)
Attend default school at baseline	0.0206	0.0148	0.0165
	(0.0131)	(0.0134)	(0.0134)
Left Montreal	0.0009	-0.0018	-0.0018
	(0.0049)	(0.0048)	(0.0048)
Left the province	0.0004	0.0017	0.0021
	(0.0036)	(0.0035)	(0.0036)
Predicted educational attainment	0.0002	0.0009	0.0006
	(0.0016)	(0.0025)	(0.0097)
Cohort fixed effects	x	x	x
Individual characteristics	x	x	x
Neighborhood (FSA) fixed effects	x	x	x
Boundary fixed effects	x	x	x

Notes: In all specifications, the control function for distance to boundary is linear and allows for different slopes on either side of the threshold. The sample includes all permanent residents, expect for the attrition variables (Left Montreal and Left the province), where all students in the database are included. All standard errors are clustered at the French primary school boundary level.

Table A5: Exposure Effects: Alternative Outcomes

Sample:	All movers		One-time movers	
	(1)	(2)	(3)	(4)
Measure of educational attainment				
No Secondary school qualification	-0.0676	-0.0648	-0.0496	-0.0496
	(0.0137)	(0.0143)	(0.0159)	(0.0165)
College enrollment (ever)	-0.0373	-0.0356	-0.0267	-0.0258
	(0.0109)	(0.0118)	(0.0134)	(0.0138)
College enrollment by 17	-0.0412	-0.0382	-0.0408	-0.0389
	(0.0081)	(0.0081)	(0.0112)	(0.0114)
College degree	-0.0407	-0.0389	-0.0336	-0.0321
	(0.0087)	(0.0090)	(0.0110)	(0.0111)
University enrollment by 19	-0.0395	-0.0381	-0.0454	-0.0442
	(0.0110)	(0.0112)	(0.0160)	(0.0163)
Bachelor degree or more	-0.0374	-0.0363	-0.0261	-0.0258
	(0.0129)	(0.0130)	(0.0181)	(0.0182)
Expected earnings on basis of	-0.0454	-0.0433	-0.0411	-0.0397
level of education	(0.0087)	(0.0093)	(0.0097)	(0.0097)
Expected earnings on basis of	-0.0406	-0.0391	-0.0334	-0.0324
level and field of education	(0.0085)	(0.0090)	(0.0106)	(0.0105)
Cohort fixed effects	x	x	x	x
Individual characteristics	x	x	x	x
Age at move fixed effects	x	x	x	x
Origin-by-destination fixed effects	x	x	x	x
Only moved once			x	x
Times in difficulty before moving		x		x
N	24316	24316	15533	15533

Notes: Note: Coefficients shown in the table are convergence rates β. In columns (2) and (4), the model includes a set of dummies for each possible value of number of times in difficulty prior to moving. Standard errors are clustered at the destination neighborhood level. Details on the measurement of outcomes are provided in the Data Appendix.

Table A6: Alternative Decomposition of Exposure Effects

Sample:	All movers			One-time movers		
	(1)	(2)	(3)	(4)	(5)	(6)
University enrollment						
			Total exposure effects			
β	-0.0424	-0.0424	-0.0424	-0.0416	-0.0416	-0.0416
	(0.0090)	(0.0090)	(0.0090)	(0.0116)	(0.0116)	(0.0116)
			School and non-school components			
β^{school}	-0.0322	-0.0321	-0.0279	-0.0322	-0.0321	-0.0283
	(0.0072)	(0.0072)	(0.0070)	(0.0095)	(0.0093)	(0.0096)
$\beta^{non\text{-}school}$	-0.0102	-0.0103	-0.0145	-0.0094	-0.0095	-0.0133
	(0.0041)	(0.0041)	(0.0058)	(0.0065)	(0.0065)	(0.0092)
Share school effects (S^{school})	76%	76%	66%	77%	77%	68%
	(0.0771)	(0.0768)	(0.1083)	(0.1304)	(0.1299)	(0.1816)
Secondary school diploma in 5 years						
			Total exposure effects			
β	-0.0421	-0.0421	-0.0421	-0.0506	-0.0506	-0.0506
	(0.0088)	(0.0088)	(0.0088)	(0.0117)	(0.0117)	(0.0117)
			School and non-school components			
β^{school}	-0.0305	-0.0286	-0.0303	-0.0397	-0.0370	-0.0385
	(0.0083)	(0.0081)	(0.0081)	(0.0109)	(0.0108)	(0.0108)
$\beta^{non\text{-}school}$	-0.0116	-0.0135	-0.0118	-0.0109	-0.0136	-0.0121
	(0.0039)	(0.0044)	(0.0038)	(0.0054)	(0.0057)	(0.0051)
Share school effects (S^{school})	72%	68%	72%	79%	73%	76%
	(0.0912)	(0.0986)	(0.0857)	(0.0989)	(0.1059)	(0.0937)
Years of education						
			Total exposure effects			
β	-0.0488	-0.0488	-0.0488	-0.0444	-0.0444	-0.0444
	(0.0088)	(0.0088)	(0.0088)	(0.0103)	(0.0103)	(0.0103)
			School and non-school components			
β^{school}	-0.0341	-0.0328	-0.0223	-0.0326	-0.0312	-0.0229
	(0.0081)	(0.0078)	(0.0085)	(0.0097)	(0.0094)	(0.0109)
$\beta^{non\text{-}school}$	-0.0147	-0.0159	-0.0265	-0.0118	-0.0132	-0.0214
	(0.0043)	(0.0045)	(0.0075)	(0.0060)	(0.0062)	(0.0101)
Share school effects (S^{school})	70%	67%	46%	73%	70%	52%
	(0.0834)	(0.0833)	(0.1386)	(0.1254)	(0.1268)	(0.2061)
Measure of school quality	$\pi\Omega_{s(n)}$	$\pi\Omega^{-i}_{s(n)}$	$\pi\Omega^{-i}_{s(n)}$	$\pi\Omega_{s(n)}$	$\pi\Omega^{-i}_{s(n)}$	$\pi\Omega^{-i}_{s(n)}$
π	1	1	RD estimate	1	1	RD estimate

Notes: Sample restricted to movers within Montreal. Standard errors are clustered at the destination FSA level, and obtained by the delta method. Parameters are defined as follow: $\beta^{school} = \beta_s\left(\frac{Var^r(\pi\Delta\Omega_{od})+Cov^r(\pi\Delta\Omega_{od},\Delta\bar{y}_{od}^{-s})}{Var^r(\Delta\bar{y}_{od})}\right)$, $\beta^{non-school} = \beta_n\left(\frac{Var^r(\Delta\bar{y}_{od}^{-s})+Cov^r(\pi\Delta\Omega_{od},\Delta\bar{y}_{od}^{-s})}{Var^r(\Delta\bar{y}_{od})}\right)$, and $S^{school} = \frac{\beta^{school}}{\beta}$. In columns (1) and (2), π is set to one. In column (3), π is given by the RD-IV estimates reported in Table 2. In columns (1) through (3), total exposure effects β correspond to estimates reported in Table 3, column (1). In columns (4) through (6), β correspond to estimates reported in Table 3, column (3).

Table A7: School Effects: Quadratic Control Function

	First-stage(s)			Reduced-form RD	RD-IV
Dependent variable:	Quality of *assigned* school at baseline ($\delta^{P}_{s(i)}$)	Quality of school *attended* at baseline ($\delta^{P}_{s(i)}$)	Childhood average school quality ($\Omega^{-i}_{s(n(i))}$)	Outcome	Outcome
	(1)	(2)	(3)	(4)	(5)
Measure of educational attainment					
			All permanent residents		
University enrollment	0.0634	0.0248	0.0293	0.0206	0.7086
	(0.0032)	(0.0028)	(0.0065)	(0.0093)	(0.2235)
Secondary school diploma in 5 years	0.0714	0.0304	0.0325	0.0351	1.0812
	(0.0036)	(0.0026)	(0.0063)	(0.0091)	(0.1891)
Years of schooling	0.2961	0.1192	0.1372	0.1098	0.8023
	(0.0146)	(0.0126)	(0.0309)	(0.0428)	(0.1914)
N	43296	43279	43291	43296	43291
			Placebo: Students in English schools		
University enrollment	0.0624	-0.0051	-0.0089	-0.0169	-
	(0.0043)	(0.0041)	(0.0106)	(0.0178)	-
Secondary school diploma in 5 years	0.0712	0.0056	0.0012	-0.0052	-
	(0.0054)	(0.0028)	(0.0077)	(0.0135)	-
Years of schooling	0.281	0.0042	-0.0075	-0.0438	-
	(0.0189)	(0.0167)	(0.0478)	(0.0770)	-
N	13446	13444	13444	13446	
Cohort fixed effects	x	x	x	x	x
Individual characteristics	x	x	x	x	x
Neighborhood (FSA) fixed effects	x	x	x	x	x
Boundary fixed effects	x	x	x	x	x

Notes: This table reports RD estimates. In columns (1) and (2), primary school quality is measured using the fixed effects, $\delta^{P}_{s(i)}$, estimated in Section (IV.A). In column (3), the dependent variable is childhood average school quality $\Omega^{-i}_{s(n(i))}$. Column (5) reports 2SLS estimates of equations (9) and (10). In all specifications, the control function for distance to boundary is quadratic and allows for different functions on either side of the threshold. In the first three rows, the sample includes all permanent residents. In the last three rows, only permanent residents enrolled in English schools are included. All standard errors are clustered at the French primary school boundary level.

Table A8: School Effects: Triangular Kernel Control Function

	First-stage(s)			Reduced-form RD	RD-IV
Dependent variable:	Quality of *assigned* school at baseline ($\delta^P_{s(i)}$)	Quality of school *attended* at baseline ($\delta^P_{s(i)}$)	Childhood average school quality ($\Omega^{-i}_{s(n(i))}$)	Outcome	Outcome
	(1)	(2)	(3)	(4)	(5)
Measure of educational attainment					
			All permanent residents		
University enrollment	0.0632	0.0245	0.0315	0.0253	0.8081
	(0.0032)	(0.0027)	(0.0064)	(0.0086)	(0.1774)
Secondary school diploma in 5 years	0.0715	0.0298	0.033	0.0344	1.0459
	(0.0036)	(0.0025)	(0.0061)	(0.0086)	(0.1700)
Years of schooling	0.2946	0.1162	0.1456	0.1129	0.7779
	(0.0146)	(0.0121)	(0.0297)	(0.0398)	(0.1661)
N	43296	43279	43291	43296	43291
			Placebo: Students in English schools		
University enrollment	0.063	-0.0025	-0.0036	-0.0110	-
	(0.0043)	(0.0041)	(0.0098)	(0.0160)	-
Secondary school diploma in 5 years	0.0721	0.0037	0.0005	-0.0084	-
	(0.0057)	(0.0026)	(0.0071)	(0.0118)	-
Years of schooling	0.2836	0.0053	0.0044	-0.0471	-
	(0.0198)	(0.0156)	(0.0433)	(0.0658)	-
N	13446	13444	13444	13446	
Cohort fixed effects	x	x	x	x	x
Individual characteristics	x	x	x	x	x
Neighborhood (FSA) fixed effects	x	x	x	x	x
Boundary fixed effects	x	x	x	x	x

Notes: This table reports RD estimates. In columns (1) and (2), primary school quality is measured using the fixed effects, $\delta^P_{s(i)}$, estimated in Section (IV.A). In column (3), the dependent variable is childhood average school quality $\Omega^{-i}_{s(n(i))}$. Column (5) reports 2SLS estimates of equations (9) and (10). In all specifications, the control function for distance to boundary is a triangular kernel and allows for different functions on either side of the threshold. In the first three rows, the sample includes all permanent residents. In the last three rows, only permanent residents enrolled in English schools are included. All standard errors are clustered at the French primary school boundary level.

Table A9: School Effects: CCT Optimal Bandwidths

	Parameters		Estimates		
	CCT Bandwith (in meters)	N	First-stage	Reduced-form	RD-IV
	(1)	(2)	(3)	(4)	(5)
Measure of educational attainment					
University enrollment	334.5	29304	0.0243	0.0192	0.788
			(0.00708)	(0.0126)	(0.404)
Secondary school diploma in 5 years	379.4	31321	0.0279	0.0271	0.972
			(0.00710)	(0.0121)	(0.327)
Years of schooling	304.5	27799	0.116	0.0977	0.846
			(0.0361)	(0.0604)	(0.379)
Cohort fixed effects			x	x	x
Individual characteristics			x	x	x
Neighborhood (FSA) fixed effects			x	x	x
Boundary fixed effects			x	x	x

Notes: This table reports RD estimates using optimal bandwidths based on Calonico, Cattaneo and Titiunik (2014). Optimal bandwidths are calculated on residualized outcomes, treating the reduced-form relationship as a sharp RD in which the binary treatment corresponds to being assigned a better default option. In all specifications, the control function for distance to boundary is a uniform kernel and allows for different functions on either side of the threshold. All standard errors are clustered at the French primary school boundary level.

Table A10: School Effects: Re-weighted Sample

	First-stage	Reduced-form	RD-IV
Dependent variable:	Childhood average school quality ($\Omega^{-i}_{s(n(i))}$)	Outcome	Outcome
	(1)	(2)	(3)
Measure of educational attainment			
University enrollment	0.0339	0.026	0.7706
	(0.0061)	(0.0083)	(0.1670)
Secondary school diploma in 5 years	0.0345	0.0351	1.0189
	(0.0059)	(0.0089)	(0.1837)
Years of schooling	0.1557	0.1095	0.705
	(0.0286)	(0.0376)	(0.1580)
N	43287	43292	43287
Cohort fixed effects	x	x	x
Individual characteristics	x	x	x
Neighborhood (FSA) fixed effects	x	x	x
Boundary fixed effects	x	x	x

Notes: This table reports RD estimates, where the sample of permanent residents is re-weighted so that its distribution of covariates matches that of the movers' sample. The matching weights are obtained using nearest-neighbor matching (5 nearest neighbors) with the Stata command `kmatch` (Jann, 2017). In all specifications, the control function for distance to boundary is linear and allows for different slopes on either side of the threshold. All standard errors are clustered at the French primary school boundary level.

Table A11: Robustness: Exposure-weighted Neighborhood Quality for Multiple-times Movers

Sample:	All movers			
	(1)	(2)	(3)	(4)
Measure of educational attainment				
University enrollment	-0.0534	-0.0512	-0.0498	-0.0484
	(0.0096)	(0.0096)	(0.0096)	(0.0097)
Secondary school diploma in 5 years	-0.0515	-0.0483	-0.0467	-0.0451
	(0.0089)	(0.0089)	(0.0090)	(0.0092)
Years of schooling	-0.0607	-0.0582	-0.0563	-0.055
	(0.0095)	(0.0099)	(0.0098)	(0.0102)
N	24316	24316	24191	24191
Cohort fixed effects	x	x	x	x
Individual characteristics	x	x	x	x
Age at move fixed effects	x	x	x	x
Origin-by-destination fixed effects	x	x	x	x
Number of moves fixed effects	x	x	x	x
Other locations controls			x	x
Times in difficulty before moving to *d*		x		x

Notes: Coefficients shown in the table are convergence rates β. The change in neighborhood quality is measured by $\bar{y}_d - E(\bar{y}_n|premove)$, where $E(\bar{y}_n|premove)$ is the exposure-weighted average neighborhood quality for all locations in which student i resided before moving to the final destination d. Note that $\bar{y}_d - E(\bar{y}_n|premove) = \Delta\bar{y}_{od}$ for one-time movers. All specifications include dummies for number of moves before the age of 15. In columns (3) and (4), fixed effects for the second and third location (prior to moving to area d), as well as for the age at which these moves occurred, are included (the omitted categories are no second/third location). Standard errors are clustered at the final destination neighborhood level.

Table A12: Robustness: 6-digit Postal Code Fixed Effects

Sample:	All movers			One-time movers		
	(1)	(2)	(3)	(4)	(5)	(6)
Measure of educational attainment						
University enrollment	-0.0424	-0.0412	-0.0403	-0.0416	-0.0408	-0.0538
	(0.0090)	(0.0092)	(0.0117)	(0.0116)	(0.0115)	(0.0214)
Secondary school diploma in 5 years	-0.0421	-0.0402	-0.0443	-0.0506	-0.0502	-0.0301
	(0.0088)	(0.0088)	(0.0151)	(0.0117)	(0.0117)	(0.0227)
Years of schooling	-0.0488	-0.0471	-0.0462	-0.0444	-0.0435	-0.0409
	(0.0088)	(0.0094)	(0.0116)	(0.0103)	(0.0102)	(0.0189)
Cohort fixed effects	x	x	x	x	x	x
Individual characteristics	x	x	x	x	x	x
Age at move fixed effects	x	x	x	x	x	x
Origin-by-destination fixed effects	x	x	x	x	x	x
Only moved once				x	x	x
Times in difficulty before moving		x	x		x	x
Destination 6-digit postal code fixed effects			x			x
N	24316	24316	16525	15533	15533	8856

Notes: Coefficients shown in the table are convergence rates β. Individual characteristics include gender, immigrant status, allophone status, born in Canada but outside Quebec, English spoken at home, day care use at baseline, "in difficulty" status at baseline, handicapped status. In columns (2) and (4), the model includes a set of dummies for each possible value of number of times in difficulty prior to moving. Columns (3) and (6) control for 6-digit postal code fixed effects at age 15. Standard errors are clustered at the destination neighborhood level.

Table A13: Balancing Tests for Movers

Outcome of permanent residents:	Years of Education		DES in 5 years		University Enrollment	
	(1)	(2)	(3)	(4)	(5)	(6)
Covariates						
Gender	0.0038	0.0034	0.0175	0.0174	0.0187	0.0162
	(0.0019)	(0.0029)	(0.0093)	(0.0136)	(0.0095)	(0.0138)
Speaks English at Home	-0.0024	-0.0001	-0.0149	-0.0035	-0.0121	-0.0034
	(0.0016)	(0.0021)	(0.0076)	(0.0107)	(0.0084)	(0.0113)
Speaks neither French nor English at Home	-0.0006	-0.0010	0.0033	0.0036	-0.0030	-0.0044
	(0.0015)	(0.0020)	(0.0077)	(0.0093)	(0.0081)	(0.0107)
Immigrant	-0.0032	-0.0045	-0.0054	-0.0104	-0.0175	-0.0246
	(0.0013)	(0.0019)	(0.0069)	(0.0095)	(0.0065)	(0.0098)
Handicapped	-0.0006	-0.0008	-0.0039	-0.0037	-0.0036	-0.0049
	(0.0006)	(0.0008)	(0.0027)	(0.0039)	(0.0027)	(0.0036)
Use Day Care at baseline	0.0003	-0.0005	-0.0004	-0.0021	0.0026	-0.0016
	(0.0014)	(0.0016)	(0.0069)	(0.0081)	(0.0066)	(0.0078)
In difficulty at baseline	0.0017	0.0016	0.0066	0.0052	0.0078	0.0085
	(0.0007)	(0.0009)	(0.0036)	(0.0045)	(0.0035)	(0.0043)
Times in difficulty pre-move	0.0098	0.0078	0.0313	0.0125	0.0365	0.0391
	(0.0081)	(0.0083)	(0.0394)	(0.0399)	(0.0394)	(0.0407)
Cohort fixed effects	x	x	x	x	x	x
Age at move fixed effects	x	x	x	x	x	x
Origin-by-Destination fixed effects	x	x	x	x	x	x
One-time movers only		x		x		x
N	24316	15533	24316	15533	24316	15533

Notes: In Columns (1) and (2), $\Delta \bar{y}_{od}$ is measured using years of education. In columns (3) and (4), fractions of students finishing secondary school in 5 years are used, and in columns (5) and (6), university enrollment rates are. Standard errors are clustered at the destination neighborhood level.

Table A14: Robustness to Time-varying Observables

	(1)	(2)	(3)	(4)	(5)	(6)
Measure of educational attainment						
Secondary school diploma in 5 years	-0.0392	-0.0437	-0.0423	-0.0370	-0.0380	-0.0318
	(0.0101)	(0.0087)	(0.0093)	(0.0100)	(0.0105)	(0.0111)
University enrollment	-0.0373	-0.0436	-0.0392	-0.0350	-0.0399	-0.0332
	(0.0107)	(0.0095)	(0.0097)	(0.0110)	(0.0107)	(0.0117)
Years of schooling	-0.0435	-0.0484	-0.0437	-0.0422	-0.0457	-0.0376
	(0.0097)	(0.0090)	(0.0090)	(0.0103)	(0.0095)	(0.0105)
Time-varying controls						
Income	x					x
Percent low-income		x				x
Dwelling value			x			x
Percent lone family				x		x
Percent with college					x	x
Cohort fixed effects	x	x	x	x	x	x
Individual characteristics	x	x	x	x	x	x
Age at move fixed effects	x	x	x	x	x	x
Origin-by-destination fixed effects	x	x	x	x	x	x
N	22735	22735	22735	22735	22735	22735

Notes: Time-varying controls are differences in census tract characteristics around the time of the move. The model includes both the main effect of these controls as well as their interaction with age-at-move. Each column includes a different set of observable time-varying variables. Standard errors are clustered at the destination neighborhood level.

Table A15: Heterogeneous Exposure Effects

Heterogeneity by:	Gender		Language at school		Moves to	
	Boys	Girls	French	English	Better FSA	Worse FSA
	(1)	(2)	(3)	(4)	(5)	(6)
Measure of educational attainment						
Secondary school diploma in 5 years	-0.0440	-0.0476	-0.0473	-0.0449	-0.0385	-0.0115
	(0.0110)	(0.0140)	(0.0114)	(0.0205)	(0.0137)	(0.0235)
University enrollment	-0.0321	-0.0571	-0.0385	-0.0390	-0.0398	-0.0536
	(0.0123)	(0.0128)	(0.0107)	(0.0207)	(0.0192)	(0.0207)
Years of schooling	-0.0425	-0.0587	-0.0485	-0.0525	-0.0257	-0.0520
	(0.0119)	(0.0127)	(0.0124)	(0.0173)	(0.0151)	(0.0192)
Cohort fixed effects	x	x	x	x	x	x
Individual characteristics	x	x	x	x	x	x
Age at move fixed effects	x	x	x	x	x	x
Origin-by-destination fixed effects	x	x	x	x	x	x
N	11600	11283	17479	5832	10981	13335

Notes: Column (1) includes only boys and column (2) restricts the sample to girls. In columns (3) and (4), regressions are run separately by language of instruction at age 15. In column (5), the sample is restricted to movers for which $\Delta \bar{y}_{od} > 0$ and column (6) is restricted to cases where $\Delta \bar{y}_{od} < 0$. Standard errors are clustered at the destination neighborhood level.

Table A16: Decomposition of Exposure Effects - Empirical Bayes Estimates

Outcome:	University enrollment		DES in 5 years		Years of education	
	(1)	(2)	(3)	(4)	(5)	(6)
			Total exposure effects			
One-step β (eq. (13))	-0.0451	-0.0451	-0.0460	-0.0460	-0.0522	-0.0522
	(0.0097)	(0.0097)	(0.0098)	(0.0098)	(0.0094)	(0.0094)
Two-step β (eqs. (14)-(15))	-0.0492	-0.0492	-0.0435	-0.0435	-0.0545	-0.0545
	(0.0105)	(0.0105)	(0.0102)	(0.0102)	(0.0103)	(0.0103)
			School and non-school components			
β^{school}	-0.0411	-0.0386	-0.0387	-0.0409	-0.0442	-0.0363
	(0.0090)	(0.0085)	(0.0102)	(0.0108)	(0.0094)	(0.0077)
$\beta^{non\text{-}school}$	-0.0080	-0.0106	-0.0048	-0.0026	-0.0102	-0.0182
	(0.0031)	(0.0033)	(0.0018)	(0.0019)	(0.0026)	(0.0033)
Share school effects (S^{school})	84%	78%	89%	94%	81%	67%
	(0.0507)	(0.0476)	(0.0461)	(0.0486)	(0.0453)	(0.0372)
Measure of school quality	$\pi\Omega^{EB}_n$	$\pi\Omega^{EB}_n$	$\pi\Omega^{EB}_n$	$\pi\Omega^{EB}_n$	$\pi\Omega^{EB}_n$	$\pi\Omega^{EB}_n$
π	1	RD estimate	1	RD estimate	1	RD estimate

Notes: Sample restricted to movers within Montreal. To shrink estimates of $\bar{y}_n$, $\Omega_{s(n(i))}$ and Λ_n, I first calculate standard errors for each school and neighborhood fixed effect using bootstrap resampling (100 samples with replacement, clustering within primary school-secondary school-FSA cells). I then shrink estimates toward their means using the empirical Bayes procedure described in Chandra et al. (2016). Note that the empirical Bayes measures of school effects are shrunk estimates of the sum of primary and secondary school effects. Standard errors on the convergence rates and school shares are clustered at the destination FSA level, and obtained by the delta method. In the first row, β is estimated via equation (11) using $m_i \times \Delta\bar{y}^{EB}_{od}$ as the main regressor, where $\Delta\bar{y}^{EB}_{od} = \bar{y}^{EB}_d - \bar{y}^{EB}_o$ and $\bar{y}^{EB}_n$ are empirical Bayes shrunk estimates of $\bar{y}_n$. In the second row, β is estimated via equations (12) and (13). For consistency, the two-step full convergence rate is used to calculate the school share. In column (1), π is set to one. In column (2), π is estimated using $\Omega^{EB}_{s(n(i))}$ in the RD equations (9) and (10), and $\Delta\bar{y}^{-s,EB}_{od} = (\Omega^{EB}_n + \Lambda^{EB}_n) - \pi\Omega^{EB}_n$ to ensure that the school and non-school shares sum to one.

B Data Appendix

Measurement of outcomes Different levels of education are governed by different departments of the Ministry of Education. Each department keeps separate student records in different formats, but these files can be matched using unique student IDs. Researchers interested in using these data must first submit a research protocol to the Ministry and file a data access request through the *Commission d'accès à l'information*.

Primary and secondary school levels, as well as vocational studies, are governed by the same department. These records notably include any secondary school degree or qualification received, vocational degrees awarded, and the year these degrees were earned. For vocational degrees, the subject is also recorded. From these files, I create an indicator variable for obtaining a secondary school diploma (*DES*) within 5 years of starting secondary school (i.e. the year a student is first observed in grade 7). Note that a student may have been held back in primary school and still obtain a secondary school diploma on time.

The College department records the year a student was first enrolled in any collegial program in Quebec, as well as the program and the institution of that first registration. If a college degree is awarded, the program in which the degree was awarded is recorded (e.g. pre-university degree in Natural Sciences). The exact date the degree was earned is not recorded, however. The files instead indicate whether the degree was completed either (a) on time, (b) less than 2 years after expected duration, or (c) more than 2 years after the expected duration. There is a further caveat: degree completion is only recorded for students who first enrolled in a "normal" college program (*DEC*). For example, degree completion is not recorded for students who first enrolled in a transition program. I use these files to create indicators of college enrollment and college completion. I also approximate the year of completion using the coarse information on time to completion.

The University department records enrollment separately by semester (Fall, Winter and Summer). For each semester, if a student is enrolled in a Quebec university, the number of credits taken, the institution and the field of study are recorded. A separate file is kept for degrees awarded. This file includes the year a degree is awarded, the granting institution, and the type of degree (bachelor, masters, doctoral, 1-year diploma, etc.). With these files, I notably create an indicator of university enrollment and one for bachelor degree completion.

Combining information from all three departments, I then calculate each student's highest level of education. The categories I consider are:

- No secondary school diploma or qualification
- Secondary school diploma (DES)
- Secondary school qualifications
- Vocational degree (DEP)
- Some collegial, started in "normal" program, no degree yet
- Some collegial, did not start in "normal" program
- Pre-university college degree
- Technical college degree

- Other college degree (includes 1-year degrees)
- Some university, no degree yet
- 1-year university diploma
- Bachelor degree or higher

I also calculate each student's number of years of education. Note that this variable might vary within the categories listed above. For instance, someone who dropped out in grade 9 has 9 years of education, while someone who dropped out in grade 10 has 10. Someone who took 13 years of primary/secondary schooling to obtain a DES and has no further schooling is coded as having 11 years of education (i.e. the normal time it takes to get a DES). Students who were in university for one year and then dropped out have 14 years of education (11 for primary+secondary school, 2 for college, and 1 in university), while those who stayed in university for two years before dropping out have 15 years of education. I top code the number of years of education at 16 (the time it takes to obtain a bachelor degree), however, to avoid my results being driven by outliers. For instance, I do observe a few hundreds students with 19 years of education or more (i.e. people from earlier cohorts in master and PhD programs). The number of years of education therefore incorporates information on multiple margins, e.g. retention in university, college enrollment, vocational studies after secondary school, drop out behavior, etc.

Finally, I create measures of expected earnings. To do so, I calculate earnings percentile ranks (in the national earnings distribution) for all workers aged 30-44 in the Public Use Microdata File of the 2006 Canadian Census, separately by age-group. I then calculate the mean earnings rank for each category of highest level of education, as well as for all possible combinations of level-of-education and field-of-study. Finally, I assign to each student in my data the mean earnings rank associated with her level of education in the 2006 Canadian Census (or combination of highest level of education and field of study). Note that students in the 1995 cohort normally finished secondary school in 2005-2006, meaning that 2006 is the year they were making their decision to pursue a post-secondary education.

Measurement of $\Omega^{-i}_{s(n(i))}$ Equation (8) simultaneously includes primary and secondary school fixed effects. This yields one fixed effect for each school in Montreal. Note that students attending a given secondary school need not have attended the same primary school – secondary schools do not nest primary schools.[41]

For each student, I then create a leave-self-out measure for both primary and secondary schools. For instance, for student i and primary school s (which student i attended), I calculate $\delta^{-i,P}_s = \frac{\delta^P_{s(i)} N_s - \tilde{y}_i}{N_s - 1}$, where N_s is the number of permanent residents who attend school s and $\tilde{y}_i = y_i - \bar{y}$ is the deviation of student i's outcome from the sample mean. Student i's outcome must be first re-centered around the sample mean because fixed effects are normalized to have a mean of zero.[42]

Then, I assign the relevant leave-self-out measure to each student-year observation. For years in which a student is in a primary school other than the one he was attending at baseline, no leave-self-out adjustment is necessary since that student was not in that school during the year on which the fixed effect estimation is based. I then take the student-level average of $\delta^{-i,P}_s$ over all primary school years, and similarly calculate a student-level average of $\delta^{-i,S}_s$ for secondary school years. The childhood school quality measure $\Omega^{-i}_{s(n(i))}$ is

[41] Default French primary schools do feed into default secondary schools. But with open enrollment, and the large number of private secondary schools, the connection between local primary and secondary schools is weak.

[42] Jackknife estimates of school fixed effects δ^{-i}_s , in which one regression is ran for each observation, are almost perfectly correlated (0.99) with my hand-calculated leave-self-out measures.

then the simple sum of these two averages. Note this averaging over primary/secondary school years only matters for permanent residents who have switched school at some point. For the majority of students who only attended one primary and one secondary school, the averaging is redundant, and it is simply the case that $\Omega^{-i}_{s(n(i))}=\delta_s^{-i,P} + \delta_s^{-i,S}$.

In unreported analyses, I use a split-sample approach in which a random half of the sample of permanent residents is used to measure school quality and the other half is used to estimate the regression-discontinuity design. Split-sample and leave-self-out measures of school quality are highly correlated (0.98), hence the results presented in this paper are very similar under the split-sample approach.

Catchment Areas To my knowledge, no electronic, geocoded version of the catchment areas that prevailed in the years 1995-2001 exists. I therefore re-constructed such maps using the following procedure.

To first generate a benchmark, the default school associated with each six-digit postal code of the Island of Montreal as of 2015 was recorded by "feeding" each of these ≈45,000 postal codes in the search engines of the websites of the three francophone schools boards. Using shapefiles for Canadian postal codes, I then created a map of all 2015 French catchment areas on the Island of Montreal, down to the six-digit postal code level.

To infer what the boundaries were in the years the cohorts of students I track started grade school, I used two additional sources of information. First, the Ministry of Education provided me temporarily with baseline enrollment data for all 100,929 students in my data set along with their six-digit postal codes (in the analytical data set, six-digit postal codes are de-identified).[43] I then mapped actual attendance patterns and compared with the 2015 boundaries. Second, I used the Internet Archives WaybackMachine (https://archive.org/web/) to document each school opening/closure that happened since 1995, and extracted old maps of catchment areas from archived versions of the school boards websites (when available). Combining all these sources of information, I deducted where the boundaries must have been drawn, and assigned the appropriate default schools to each postal code by hand. It must be noted that for many schools, the boundaries have not changed since 1995, hence no manual re-coding was necessary. Using ArcGIS, I also calculated, for each postal code, the distance to the nearest boundary and the unique ID of that boundary. Only boundaries that do not coincide with major geographical features, such as highways or canals, were considered. Using these same sources of information, I also inputted catchment areas for English public schools. As explained in the text, however, these boundaries are not well-defined and therefore not used in the analyses.

Attrition About 8% of the total number of students who started grade 1 in Montreal had vanished from primary/secondary school educational records before turning 16. These students are excluded from the main sample used this paper. Interestingly, about 1,000 of these students did enroll in a Quebec university at some point, even though they did not graduate from secondary school in the province. Students who had left Montreal (but remained in Quebec) by the time they turned 15 are also excluded from all analyses.

For higher-education, enrollment in colleges and universities outside the province is not comprised in my dataset. As a result, I may wrongly infer that some students in my main sample never attended college, when in fact they did out-of-province. However, this phenomenon likely only affects a very small proportion of my sample. A few factors provide strong incentives for college and university students to remain in the province, at least for their undergraduate studies. Firstly, tuition fees in Quebec are the lowest in Canada. Secondly, the discrepancies between Quebec's and other North American educational systems generate important

[43] This first data delivery contained only two variables: school attended (name and code) and postal code of residence. For confidentiality reasons, this file had to be destroyed before the analytical files could be transferred to me.

timing issues in meeting college requirements. For instance, at the end of secondary school, students in Quebec only have 11 years of school, rather than 12. Finally, there is a language barrier for the large majority of students who went to primary and secondary school in French.

To assess the possible magnitude of this measurement issue, I use data from the loans and bursaries records of the Ministry of Education. For each year between 1995-1996 and 2014-2015, I was given a series of indicator variables that flag whether student i in my sample was receiving loans or bursaries in year t. Students who resided in Quebec in childhood but go abroad for college are still eligible for loans and bursaries from the Quebec government. Since at the time of enrolling in a foreign college the student's permanent address is often still a Quebec one, it is easier for them to take up loans from Quebec than from another province. I can therefore check the proportion of students who take up students loans while not being enrolled in any postsecondary institution in Quebec to assess the size of the phenomenon. Under this method, I find that about 1% of my sample attended a higher education institution outside the province at some point (many of which also attended a college or a university in Quebec before doing so out-of-province). Finally, it is worth noting that any mis-measurement of educational attainment due to students leaving the province would plausibly lead me to *underestimate* differences across schools and neighborhoods. Students studying abroad, where tuition is much more expensive, are arguably from higher-SES backgrounds, leading me to underestimate educational attainment in places where it is the highest.

C Mathematical Appendix

C.1 Interpretation of π: Example

Suppose we did observe μ_n and $\tilde{\psi}_{s(n(i))}$, and ran a simple cross-sectional regression of $y^{PR}_{n(i)}$ on both these variables for the subsample of permanent residents. The regression equation would take the following form:

$$y^{PR}_{n(i)} = \alpha_n \mu_n + \alpha_s \tilde{\psi}_{s(n(i))} + \epsilon_i. \tag{17}$$

The OLS estimate of α_s is equal to $A\omega + A\rho_s$, where ρ_s corresponds to the omitted variable bias. Alternatively, ρ_s is a partial regression coefficient in a linear projection of family inputs $\tilde{\theta}_i$ onto μ_n and $\tilde{\psi}_{s(n(i))}$. Then, $\Omega_{s(n(i))} = \alpha_s \tilde{\psi}_{s(n(i))}$, $\pi = \frac{\omega}{\omega+\rho_s}$ and $\pi\Omega_{s(n(i))} = A\omega\tilde{\psi}_{s(n(i))}$.

Now, consider the feasible regression of $y^{PR}_{n(i)}$ on $\Omega_{s(n(i))}$ as well as on a set of neighborhood fixed effects, and let $\dot{\Omega}_{s(n(i))}$ denote the residuals from a regression of $\Omega_{s(n(i))}$ on neighborhood fixed effects. The OLS estimate of the coefficient on $\Omega_{s(n(i))}$ is

$$\begin{aligned}
\frac{Cov\left(y^{PR}_{n(i)}, \dot{\Omega}_{s(n(i))}\right)}{Var(\dot{\Omega}_{s(n(i))})} &= \underbrace{\frac{Cov\left(A\lambda\mu_n, \dot{\Omega}_{s(n(i))}\right)}{Var(\dot{\Omega}_{s(n(i))})}}_{=0 \text{ (fixed effects)}} + \frac{Cov\left(A\omega\tilde{\psi}_{s(n(i))}, \dot{\Omega}_{s(n(i))}\right)}{Var(\dot{\Omega}_{s(n(i))})} + \frac{Cov\left(A\tilde{\theta}_i, \dot{\Omega}_{s(n(i))}\right)}{Var(\dot{\Omega}_{s(n(i))})} \\
&= \frac{A\omega}{\alpha_s} + \frac{A\rho_s}{\alpha_s} = 1.
\end{aligned}$$

Note that $\Omega_{s(n(i))}$ is a measure of predicted gains (estimated school effects), while $A\omega\tilde{\psi}_{s(n(i))}$ corresponds to true gains (true school effects). Now consider an experimental sample of permanent residents that is randomly assigned to schools and neighborhoods. Their outcomes are given by $y^E_i = A\left[\lambda\mu_{n(i)} + \omega\tilde{\psi}_{s(n(i))} + \nu_i\right]$, where ν_i is uncorrelated with $\mu_{n(i)}$ and $\tilde{\psi}_{s(n(i))}$ by virtue of random assignment. Consider a regression of y^E_i on a measure of $\Omega_{s(n(i))}$ constructed using an external, non-experimental sample. The OLS coefficient obtained by regressing y^E_i on $\Omega_{s(n(i))}$ and a set of neighborhood fixed effects is

$$\begin{aligned}
\frac{Cov(y^E_i, \dot{\Omega}_{s(n(i))})}{Var(\dot{\Omega}_{s(n(i))})} &= \frac{Cov\left(A\left[\lambda\mu_{n(i)} + \omega\tilde{\psi}_{s(n(i))} + \nu_i\right], \dot{\Omega}\right)}{Var(\dot{\Omega})} \\
&= \underbrace{\frac{Cov\left(A\lambda\mu_{n(i)}, \dot{\Omega}\right)}{Var(\dot{\Omega})}}_{=0 \text{ (fixed effects)}} + \frac{Cov\left(A\omega\tilde{\psi}_{s(n(i))}, \dot{\Omega}\right)}{Var(\dot{\Omega})} + \underbrace{\frac{Cov\left(\nu_i, \dot{\Omega}\right)}{Var(\dot{\Omega})}}_{=0 \text{ (randomization)}} \\
&= \frac{Cov\left(A\omega\tilde{\psi}_{s(n(i))}, \dot{\Omega}\right)}{Var(\dot{\Omega})} = \frac{A\omega}{\alpha_s} = \frac{\omega}{\omega+\rho_s} = \pi.
\end{aligned}$$

The coefficient π is therefore the ratio of the causal effect of attending a better school over total school variation (the causal effect plus the sorting component). In the language of Chetty, Friedman and Rockoff (2014*a*), π is the relationship between true school effects and estimated school effects. It follows that $1 - \pi = \frac{\rho_s}{\omega+\rho_s}$ is the amount of forecast bias in $\Omega_{s(n(i))}$.

Without such an experimental sample, one can still estimate the amount of forecast bias using a valid instrumental variable Z_i that shifts $\tilde{\psi}_{s(n(i))}$ but is otherwise orthogonal to parental inputs $\tilde{\theta}_i$. The IV estimate of the coefficient on $\Omega_{s(n(i))}$ in a regression of $y^{PR}_{n(i)}$ on $\Omega_{s(n(i))}$ as well as on a set of neighborhood fixed effects is

$$\frac{Cov(y^{PR}_{n(i)}, \dot{Z}_i)}{Cov(\Omega_{s(n(i))}, \dot{Z}_i)} = \underbrace{\frac{Cov\left(A\lambda\mu_n, \dot{Z}_i\right)}{Cov(\Omega_{s(n(i))}, \dot{Z}_i)}}_{=0 \text{ (fixed effects)}} + \frac{Cov\left(A\omega\tilde{\psi}_{s(n(i))}, \dot{Z}_i\right)}{Cov(\Omega_{s(n(i))}, \dot{Z}_i)} + \underbrace{\frac{Cov\left(A\tilde{\theta}_i, \dot{Z}_i\right)}{Cov(\Omega_{s(n(i))}, \dot{Z}_i)}}_{=0 \text{ (exclusion restriction)}}$$

$$= A\omega\frac{Cov\left(\tilde{\psi}_{s(n(i))}, \dot{Z}_i\right)}{Cov(\Omega_{s(n(i))}, \dot{Z}_i)} = \frac{A\omega}{\alpha_s}\frac{Cov\left(\Omega_{s(n(i))}, \dot{Z}_i\right)}{Cov(\Omega_{s(n(i))}, \dot{Z}_i)} = \frac{\omega}{\omega+\rho_s} = \pi$$

where $\dot{Z}_i$ denotes the residuals from a regression of Z_i the neighborhood fixed effects.

C.2 Interpretation of reduced-form coefficients β_s and β_n

For ease of exposition, let the school compliance factor be constant across students: $c_{i(o,d)} = c\forall i$, and let ξ denote the relationship between estimated neighborhood (non-school) effects Λ_n and true effects $A\lambda\mu_n$. Total exposure effects simplify to

$$\begin{aligned} e_i(o,d) =& \frac{1}{A}c\pi\Delta\Omega_{od} + \frac{1}{A}\Delta\bar{y}^{-s}_{od} - \left(\bar{\theta}^{PR}_d - \bar{\theta}^{PR}_o\right). \\ =& \frac{1}{A}\left[c - \xi\frac{1-\pi}{\pi}\right]\pi\Delta\Omega_{od} + \frac{1}{A}\left[\xi\right]\Delta\bar{y}^{-s}_{od} \end{aligned} \tag{18}$$

which implies that $\beta_s = \frac{1}{A}\left(c - \xi\frac{1-\pi}{\pi}\right)$ and $\beta_n = \frac{1}{A}\xi$. The first coefficient, β_s, is increasing in the compliance rate c, but decreasing in absolute $(1-\pi)$ and relative $\frac{\xi}{\pi}$ sorting into schools. The two coefficients β_s and β_n are equal when $c = 1$ and $\xi = \pi$. As a result, the school share S^{school} is decreasing in the degree of sorting into schools (i.e. it converges to zero as $\rho_s \to \infty$), and the sorting term $\left(\bar{\theta}^{PR}_d - \bar{\theta}^{PR}_o\right)$ is allocated to the school and non-school components in proportion of the relative amount of sorting in each dimension. For example, assume full compliance $c = 1$ and let the variance of true school effects be the same as the variance of true neighborhood (non-school) effects: $Var(A\omega\Delta\psi_{od}) = Var(A\lambda\Delta\mu_{od})$. In this case, $\beta_s Var(A\omega\Delta\psi_{od}) > \beta_n Var(A\lambda\Delta\mu_{od})$ if and only if $\pi > \xi$. Similarly, if the degree of sorting is the same in both dimensions $\xi = \pi$ (maintaining the assumption of full compliance), it follows that $\beta_s Var(A\omega\Delta\psi_{od}) > \beta_n Var(A\lambda\Delta\mu_{od})$ if and only if $Var(A\omega\Delta\psi_{od}) > Var(A\lambda\Delta\mu_{od})$. Note that if movers never switch schools ($c = 0$), β_s could be negative. Intuitively, for given absolute gains due to non-school factors (the numerator of the convergence rate), larger differences in school effects across locations $\pi\Delta\Omega_{od}$ imply a wider gap in educational attainment $\Delta\bar{y}_{od}$ (the denominator of the convergence rate). The convergence rate must therefore be decreasing in the relative importance of school effects if movers don't switch schools.

C.3 Decomposition equation

For ease of exposition, ignore the conditioning variables and fixed effects included in equations (11) and (12), and set $\pi = 1$ so that $(\Delta\bar{y}_{od} - \pi\Delta\Omega_{od}) = \Delta\Lambda_{od}$. Let $m\widehat{\Delta\Omega}_{od}$ denote the residuals of a regression of $m\Delta\Omega_{od}$ on $m\Delta\Lambda_{od}$: $m\widehat{\Delta\Omega}_{od} = m\Delta\Omega_{od} - \frac{Cov(m\Delta\Omega_{od}, m\Delta\Lambda_{od})}{Var(m\Delta\Lambda_{od})}m\Delta\Lambda_{od}$. Define $m\widehat{\Delta\Lambda}_{od}$ accordingly. Then, the coefficients of the simplified horse-race regression $y_i = \beta_s(m\Delta\Omega_{od}) + \beta_n(m\Delta\Lambda_{od})$ are

$$\beta_s = \frac{Cov\left(m\widehat{\Delta\Omega}_{od}, y_i\right)}{Var(m\widehat{\Delta\Omega}_{od})} \quad ; \quad \beta_n = \frac{Cov\left(m\widehat{\Delta\Lambda}_{od}, y_i\right)}{Var(m\widehat{\Delta\Lambda}_{od})}.$$

The associated full convergence rate is $\beta = \frac{Cov(y_i, m\Delta\bar{y}_{od})}{Var(m\Delta\bar{y}_{od})}$. Re-organizing

$$\begin{aligned}
Var(m\Delta\bar{y}_{od}) \times \beta &= Cov\left(m\Delta\bar{y}_{od}, y_i\right) \\
&= Cov\left(m\Delta\Omega_{od}, y_i\right) + Cov\left(m\Delta\Lambda_{od}, y_i\right) \\
&= Cov\left(m\widehat{\Delta\Omega}_{od}, y_i\right) + \frac{Cov(m\Delta\Omega_{od}, m\Delta\Lambda_{od})}{Var(m\Delta\Lambda_{od})} Cov\left(m\Delta\Lambda_{od}, y_i\right) \\
&+ Cov\left(m\widehat{\Delta\Lambda}_{od}, y_i\right) + \frac{Cov(m\Delta\Omega_{od}, m\Delta\Lambda_{od})}{Var(m\Delta\Omega_{od})} Cov\left(m\Delta\Omega_{od}, y_i\right) \\
&= \beta_s Var(m\widehat{\Delta\Omega}_{od}) + \beta_n Var\left(m\widehat{\Delta\Lambda}_{od}\right) \\
&+ Cov(m\Delta\Omega_{od}, m\Delta\Lambda_{od}) \left[\frac{Cov\left(m\Delta\Lambda_{od}, y_i\right)}{Var(m\Delta\Lambda_{od})} + \frac{Cov\left(m\Delta\Omega_{od}, y_i\right)}{Var(m\Delta\Omega_{od})}\right] \\
&= \beta_s \left[Var(m\Delta\Omega_{od}) - \frac{Cov(m\Delta\Omega_{od}, m\Delta\Lambda_{od})^2}{Var(m\Delta\Lambda_{od})}\right] \\
&+ \beta_n \left[Var(m\Delta\Lambda_{od}) - \frac{Cov(m\Delta\Omega_{od}, m\Delta\Lambda_{od})^2}{Var(m\Delta\Omega_{od})}\right] \\
&+ Cov(m\Delta\Omega_{od}, m\Delta\Lambda_{od}) \left[\frac{Cov\left(m\Delta\Lambda_{od}, y_i\right)}{Var(m\Delta\Lambda_{od})} + \frac{Cov\left(m\Delta\Omega_{od}, y_i\right)}{Var(m\Delta\Omega_{od})}\right] \\
&= \beta_s Var(m\Delta\Omega_{od}) + \beta_n Var(m\Delta\Lambda_{od}) + \left(\beta_s + \beta_n\right) Cov(m\Delta\Omega_{od}, m\Delta\Lambda_{od})
\end{aligned}$$

where the non-school share is given by

$$S^{non-school} = \frac{Cov(y_i, m\Delta\Lambda_{od})}{Cov(y_i, m\Delta\bar{y}_{od})} = \frac{1}{\beta}\left[\frac{\beta_n Var(m\Delta\Lambda_{od}) + \beta_s Cov(m\Delta\Omega_{od}, m\Delta\Lambda_{od})}{Var(m\Delta\bar{y}_{od})}\right] = \frac{\beta^{non-school}}{\beta}.$$

Note that $\beta^{non-school}$ can equivalently be estimated by regressing y_i on $m\Delta\Lambda_{od}$ (not conditioning on $m\Delta\Omega_{od}$) and rescaling the resulting coefficient by $\frac{Var(m\Delta\Lambda_{od})}{Var(m\Delta\bar{y}_{od})}$:

$$\frac{Cov(y_i, m\Delta\Lambda_{od})}{Var(m\Delta\Lambda_{od})} \frac{Var(m\Delta\Lambda_{od})}{Var(m\Delta\bar{y}_{od})} = \frac{Cov(y_i, m\Delta\Lambda_{od})}{Var(m\Delta\bar{y}_{od})} = \frac{\beta_n Var(m\Delta\Lambda_{od}) + \beta_s Cov(m\Delta\Omega_{od}, m\Delta\Lambda_{od})}{Var(m\Delta\bar{y}_{od})}.$$

The proof is similar for the more general case in which there are covariates in the estimating equations: First residualize $m\Delta\Omega_{od}$, $m\Delta\Lambda_{od}$ and $m\Delta\bar{y}_{od}$ on the relevant control variables and fixed effects. The accounting identity then becomes

$$\beta = \frac{1}{Var^r(\Delta\bar{y}_{od})}\left[\beta_s Var^r(\Delta\Omega_{od}) + \beta_n Var^r(\Delta\Lambda_{od}) + \left(\beta_s + \beta_n\right) Cov^r(\Delta\Omega_{od}, \Delta\Lambda_{od})\right]$$

where $Var^r(z)$ and $Cov(z)$ denote the variance and covariance of the residuals of $m \times z$. For instance, $Cov^r(\Delta\Omega_{od}, \Delta\Lambda_{od})$ is the covariance of the residuals of $m\Delta\Omega_{od}$ and the residuals of $m\Delta\Lambda_{od}$.

D School Effects: Specification Checks

D.1 Locally Constant School Effects

Dong and Lewbel (2015) show that in regression discontinuity settings, the change in slope at the cutoff, which they label the treatment effect derivative (TED), has implications for testing whether the LATE is locally constant. In fact, one necessary condition for not having the treatment effect vary with the running variables is that TED=0. In a local linear regression setting, this amounts to testing whether the interaction between the running variable and the treatment dummy is statistically significant. This insight implies that the coefficient on $distance_{ib} \times HighSide_b$ in equations (9) and (10) contains valuable information. In my context, this tests indicates whether it is reasonable to assume that school effects are constant with respect to distance to boundaries, and therefore to extrapolate the RD-IV effects away from boundaries.

In Table D1, the coefficients on $distance_{ib}$ and $distance_{ib} \times HighSide_b$ for my baseline specifications are reported in columns (1), (4) and (7) for university enrollment, DES in 5 years and years of education, respectively. For convenience, distance is measured in kilometers (rather than meters). The reduced-form results indicate that educational attainment is increasing with distance, plausibly because (a) households further away from boundaries live closer to the school itself, and (b) households at larger distances are generally located in suburban areas where educational attainment is relatively higher.

The coefficient on the interaction term $distance_{ib} \times HighSide_b$ is generally negative but smaller in magnitude than the main effect of distance, indicating that outcomes increase with distance on both sides of boundaries. However, both for the first-stage and the reduced-form, these interactions are small and not statistically significant. The interactions in the RD-IV second-stage equation are also small and insignificant, consistent with locally constant effects.

D.2 Linearity of π

In this section, I examine whether the assumption of a linear relationship between estimated and true school effects is reasonable. To do so, I split the sample of boundaries used in the RD analysis in two according to the size of the difference in quality between the two default French primary schools. In columns (2), (5) and (8) of Table D1, the RD-IV model is estimated on the subsample of students who live near small-gap boundaries. Similarly, in columns (3), (6) and (9), the sample is restricted to boundaries with above-average differences in quality between the two default options.

For all three outcomes, large-gap boundaries are associated with an average jump in childhood school quality $\Omega^{-i}_{s(n(i))}$ more than twice the size of the jump around small-gap boundaries. For example, the first-stage coefficient for university enrollment is 0.016 for small-gap boundaries, and 0.047 for large-gap boundaries. Similarly, the reduced-form RD coefficients for large-gap boundaries are more than twice the size of the coefficients for small-gap boundaries. As a result, the estimates of π are fairly constant across "gap size". For university enrollment, while the main RD-IV estimate is 0.85, it is only slightly smaller for small-gap boundaries (0.82) and slightly higher for large-gap boundaries (0.90). For DES in 5 years, the estimate of π is surprisingly large for small-gap boundaries (1.25), but this coefficient is not very precisely estimated (s.e. 0.55). For years of educations, π also appears to be stable across the two sets of boundaries (0.71 and 0.79).

Table D1: Constant School Effects

Outcome:	University enrollment			DES in 5 years			Years of education		
Boundaries:	All boundaries	Small differences	Large differences	All boundaries	Small differences	Large differences	All boundaries	Small differences	Large differences
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
First-Stage ($\Omega^{-i}_{s(n(i))}$)									
RD coefficient (*HighSide*)	0.0328	0.0159	0.0470	0.0337	0.0132	0.0539	0.1511	0.0916	0.2340
	(0.0065)	(0.0089)	(0.0092)	(0.0061)	(0.0077)	(0.0101)	(0.0298)	(0.0401)	(0.0454)
distance	0.0247	0.0171	0.0303	0.0326	0.0188	0.0480	0.1209	0.1379	0.0976
	(0.0118)	(0.0151)	(0.0175)	(0.0097)	(0.0096)	(0.0201)	(0.0517)	(0.0564)	(0.1065)
distance X HighSide	-0.0133	-0.0080	-0.0081	-0.0282	-0.0200	-0.0098	-0.0791	-0.0897	-0.1011
	(0.0176)	(0.0211)	(0.0292)	(0.0128)	(0.0110)	(0.0342)	(0.0705)	(0.0757)	(0.1885)
Reduced-form									
RD coefficient (*HighSide*)	0.0279	0.0131	0.0422	0.0347	0.0165	0.0528	0.1165	0.0652	0.1830
	(0.0087)	(0.0117)	(0.0129)	(0.0084)	(0.0106)	(0.0142)	(0.0390)	(0.0497)	(0.0643)
distance	0.0293	0.0207	0.0387	0.0328	0.0193	0.0481	0.1315	0.2094	0.0585
	(0.0162)	(0.0212)	(0.0257)	(0.0158)	(0.0155)	(0.0342)	(0.0715)	(0.0738)	(0.1458)
distance X HighSide	-0.0040	0.0090	-0.0254	-0.0279	-0.0200	-0.0178	-0.0759	-0.1389	-0.0817
	(0.0221)	(0.0303)	(0.0368)	(0.0176)	(0.0176)	(0.0455)	(0.0805)	(0.0812)	(0.2322)
RD-IV									
RD coefficient (*HighSide*)	0.8542	0.8242	0.9017	1.0340	1.2509	0.9837	0.7739	0.7107	0.7857
	(0.1645)	(0.4683)	(0.1740)	(0.1618)	(0.5492)	(0.1801)	(0.1575)	(0.3257)	(0.1733)
distance	0.0082	0.0067	0.0114	-0.0010	-0.0043	0.0008	0.0374	0.1109	-0.0189
	(0.0101)	(0.0146)	(0.0153)	(0.0094)	(0.0121)	(0.0188)	(0.0437)	(0.0569)	(0.0761)
distance X HighSide	0.0073	0.0155	-0.0179	0.0014	0.0051	-0.0080	-0.0139	-0.0747	-0.0005
	(0.0122)	(0.0182)	(0.0202)	(0.0098)	(0.0123)	(0.0266)	(0.0481)	(0.0511)	(0.1094)
N	43291	22928	20362	43291	22928	20362	43291	22928	20362
Cohort fixed effects	x	x	x	x	x	x	x	x	x
Individual characteristics	x	x	x	x	x	x	x	x	x
Neighborhood (FSA) fixed effects	x	x	x	x	x	x	x	x	x
Boundary fixed effects	x	x	x	x	x	x	x	x	x

Notes: Columns (1), (4) and (7) reproduce the main RD-IV results shown in Table 2. In the remaining columns, the sample is split in half according to the size of the difference in school quality $\delta^P_{s(i)}$ between the two default options split by a boundary. In columns (2), (5) and (8), only boundaries with differences below the mean are in the sample. In columns (3), (6) and (9), the sample is restricted to above-average boundaries in terms of the magnitude of the differences between the two default French primary schools. All standard errors are clustered at the French primary school boundary level.

D.3 Model-based approximation of π

As in section C.2, let the school compliance factor c_i be constant across students, and let ξ denote the relationship between estimated neighborhood (non-school) effects Λ_n and true effects $A\lambda\mu_n$. In this case, we can write

$$e_i(o,d) = \left(\frac{c\pi}{A}\right)\Delta\Omega_{od} + \left(\frac{\xi}{A}\right)\Delta\Lambda_{od}. \tag{19}$$

Now suppose we run the horse-race model in equation (12), but use $\Delta\Omega_{od}$ and $\Delta\Lambda_{od}$ as regressors. Under the above assumptions, the partial regression coefficient on $\Delta\Omega_{od}$ is equal to $\frac{c\pi}{A}$. Provided one knows the values of c and A, it is therefore possible to approximate the value of π using the movers design alone. There is one important caveat: in practice, the convergence rate is certainly not constant and likely interacts with $\Delta\Omega_{od}$ (e.g. moves on short distances might be associated with small differences in school quality as well as with low probabilities of switching school) as well as with age-at-move m_i. Hence, the *estimated* partial regression coefficient at best approximates $\frac{c\pi}{A}$.

To obtain a rough estimate of c, I conduct an event-study similar to the one described in Section V.B. For movers, I define an index of relative school quality by

$$\sigma^{\psi}_{od(i,t)} = \frac{\delta_{s(i,t)} - \bar{\delta}_{s(o,t)}}{\bar{\delta}_{s(d,t)} - \bar{\delta}_{s(o,t)}} \tag{20}$$

where $\delta_{s(i,t)}$ is the quality of the school attended by student i at time t (measured by the fixed effects estimates obtained in Section IV.A), and $\bar{\delta}_{s(n,t)}$ is the average quality of schools attended by permanent residents of FSA n at time t. The corresponding event-study results are shown in Figure A20. The index increases sharply in value right at the time of the move. While there seems to be a modest spike in the year preceding the move, this bump is very small compared to the break that occurs on impact. Pre-post differences in the index of relative school quality for one-time movers therefore provide a rough estimate of the compliance rate.[44] The estimates reported in Figure A20 for one-time movers range between 0.54 and 0.65. I therefore calibrate $c = 0.6$ below.

Model-based estimates of π are shown in Table D2. In columns (1) and (3), I calibrate the exposure period to $A = 10$ (grade 1 to grade 10), and in columns (2) and (4) I set $A = 12$ to encompass all school years, that is including kindergarten and the last year of secondary school.

Using reasonable values of A and c, I obtain estimates of π that range between 0.72 and 0.87 for university enrollment, between 0.62 and 1 for timely secondary school graduation, and between 0.72 and 0.89 for years of education. These values are quite similar to the ones obtained using the RD design, which also range between 0.6 and 1 across specifications and outcomes.

[44] Let the average index of relative school quality $\sigma^{\psi}_{od(i,t)} = \frac{\delta_{s(i,t)} - \bar{\delta}_{s(o,t)}}{\bar{\delta}_{s(d,t)} - \bar{\delta}_{s(o,t)}}$ for pre-move years, that is for years spent in the origin, be $\frac{\tilde{\delta}_{s(o,t)} - \bar{\delta}_{s(o,t)}}{\bar{\delta}_{s(d,t)} - \bar{\delta}_{s(o,t)}}$. Similarly, the average for post-move years is $\frac{\tilde{\delta}_{s(d,t)} - \bar{\delta}_{s(o,t)}}{\bar{\delta}_{s(d,t)} - \bar{\delta}_{s(o,t)}}$, and the difference between the two is $\frac{\tilde{\delta}_{s(d,t)} - \tilde{\delta}_{s(o,t)}}{\bar{\delta}_{s(d,t)} - \bar{\delta}_{s(o,t)}}$. Note that the plots that include all movers may suffer from measurement error because the index of relative school quality is likely mismeasured in pre-move years for people who move multiple time before reaching their final destination.

Table D2: Movers-based Estimates of π

Sample:	All movers		One-time movers	
	(1)	(2)	(3)	(4)
Coefficient on (m_i × $\Delta\Omega_{od}$)				
University enrollment	-0.0431	-0.0431	-0.0435	-0.0435
	(0.0097)	(0.0097)	(0.0128)	(0.0128)
Secondary school diploma in 5 years	-0.0374	-0.0374	-0.0502	-0.0502
	(0.0102)	(0.0102)	(0.0138)	(0.0138)
Years of schooling	-0.0445	-0.0445	-0.0432	-0.0432
	(0.0106)	(0.0106)	(0.0128)	(0.0128)
Calibration				
A	10	12	10	12
c	0.6	0.6	0.6	0.6
Implied value of π				
University enrollment	0.72	0.86	0.73	0.87
Secondary school diploma in 5 years	0.62	0.75	0.84	1.00
Years of schooling	0.74	0.89	0.72	0.86
Cohort fixed effects	x	x	x	x
Individual characteristics	x	x	x	x
Age at move fixed effects	x	x	x	x
Origin-by-destination fixed effects	x	x	x	x
N	24316	24316	15533	15533

Notes: In columns (1) and (2), the partial regression coefficients on $m_i \times \Delta\Omega_{od}$, say $\tilde{\beta}_s$, are estimated for all movers, and in columns (3) and (4) they are estimated on the subsample of one-time movers. The implied value of π is calculated by $\pi = -\tilde{\beta}_s A/c$.

E Level of Geography

Here, I verify that the results do not hinge on the choice of level of geography used to define neighborhoods. To do so, I consider census tracts (CT) as an alternative unit. There are about 500 different census tracts in Montreal, more than five times the number of FSAs.

By construction, there is more raw variation in educational attainment across CTs than across FSAs. However, when estimating school quality using a two-way fixed effect model (equation (8)), the amount of variation across schools, net of neighborhood fixed effects, is largely unaffected by the choice of geography. As shown in Table (E1), even when one increases the number of neighborhood units five-fold, it is still the case that there is more variation across schools than across neighborhoods.

Table E2 then presents estimates of total exposure effects when neighborhoods are defined as census tracts. In Panel A, origin-by-destination fixed effects are included. The number of effective observations in the first two columns (18,981) is considerably smaller than for the results based on FSA (24,316 movers) because there are multiple cases of origin-by-destination pairs that, at the census tract level, contain only one observation, and are therefore dropped. I therefore consider a less restrictive specification in Panel B, where origin and destination fixed effects enter separately.

In columns (1) and (2) all movers are included. The reported convergence rates tend to vary between 2% and 2.5%, considerably smaller than at the FSA level. This is likely due to the fact that sampling error is greater at smaller levels of geography (i.e. fewer permanent residents per neighborhoods), and perhaps also reflect greater sorting of permanent residents at smaller levels of geography. Census tracts may also less precisely capture all features of the community in which children live and socialize.

In columns (3) and (4), I only consider movers who, at least once, moved between FSAs, and in column (5) and (6) I only consider children who moved across census tracts but remained within the same FSA. For between-FSA moves (but still measuring neighborhood quality at the census tract level), I now find convergence rates around 2.5-3%, slightly higher than rates for all movers, but still considerably smaller than FSA-level estimates. For within-FSA moves, I find no evidence of any convergence – no estimate is statistically significant at conventional levels, and in Panel A coefficients even turn positive. For completeness, in columns (7) and (8) I restrict the sample to one-time movers, and only focus on census tract movers who moved across FSAs exactly once in columns (9) and (10). Convergence rates for these sub-groups vary between 2% and 3%.

Finally, in Table E3, I investigate whether the school share is smaller or larger when using census tracts rather than FSAs. The total exposure effects I decompose are those shown in Panel B, column (1) of Table E2. Note that for consistency, the entire procedure is re-estimated at the census tract level. In other words, the school fixed effects used in this analysis are net of census tract fixed effects, and than the RD-IV coefficient of π is similarly estimated substituting census tract fixed effects for FSA fixed effects in equations (9) (10).

Interestingly, in column (1), the school share S^{school} is slightly smaller at the census tract level than at the FSA level for university enrollment (70%) and secondary school graduation (67%), but the opposite is true for years of education (86%). The estimates that are adjusted for sorting into schools, presented in column (3), are respectively 53%, 76% and 73% for university enrollment, secondary school graduation and years of education. Overall, the estimates are neither systematically larger nor smaller than at the FSA level, but rather similar.

Two opposing forces push the school share in opposite directions when going from the FSA to the census tract level. First, because there are 5 times more census tracts than FSAs, the variance of non-school factors

$Var^r(\Delta\bar{y}_{od}^{-s})$ becomes relatively larger than the variance of school factors $Var^r(\pi\Delta\Omega_{od})$, which tends to make the school share smaller. However, because there might be more sorting of permanent residents across census tracts than across FSAs, and because census tract may only provide a noisy estimate of true neighborhood quality, β_n will likely be much smaller at the census tract level than at the FSA level (relative to β_s), making the non-school share smaller. I find that this is indeed the case. At the FSA level, it is generally the case that $\beta_n > \beta_s$, but the opposite is generally true at the census tract level.

Table E1: Variation Across Census Tracts and Schools

	Outcome					
	DES in 5 years		University enrollment		Years of education	
	(1)	(2)	(3)	(4)	(5)	(6)
Student-level standard deviation of fixed effects:						
Schools	0.261	0.255	0.248	0.235	1.172	1.123
Neighborhoods (Census Tracts)	0.152	0.068	0.159	0.081	0.734	0.328
Dependent variable summary statistics:						
Mean	0.729		0.460		13.323	
Standard deviation	[0.444]		[0.498]		[2.083]	
Fixed effects estimated						
Separately	x		x		x	
Simultaneously		x		x		x
Number of students	37,491					
Number of primary schools	435					
Number of secondary schools	211					
Number of neighborhoods	502					

Notes: Sample restricted to students who always resided in the same census tract. School fixed effects are the sum of a primary and a secondary school fixed effect. In columns (1), (3) and (5), school and neighborhood effects are respectively estimated in separate regressions. In columns (2), (4) and (6), all fixed effects are estimated simultaneously from equation (8).

Table E2: Total Exposure Effects: Moves Across Census Tracts

Sample:	All movers		Moved across FSAs at least once		Never moved across FSAs		One-time movers		Moved across FSAs exactly once	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Measure of educational attainment										
Panel A: Origin-by-destination fixed effects										
University enrollment	-0.0182	-0.0218	-0.0252	-0.0290	0.0093	0.0084	-0.0171	-0.0225	-0.0153	-0.0206
	(0.0116)	(0.0114)	(0.0137)	(0.0132)	(0.0251)	(0.0255)	(0.0191)	(0.0189)	(0.0190)	(0.0186)
Secondary school diploma in 5 years	-0.0175	-0.0200	-0.0196	-0.0222	0.0001	0.0019	-0.0182	-0.0201	-0.0300	-0.0355
	(0.0123)	(0.0115)	(0.0141)	(0.0133)	(0.0237)	(0.0228)	(0.0190)	(0.0182)	(0.0202)	(0.0196)
Years of schooling	-0.0173	-0.0219	-0.0261	-0.0315	0.0142	0.0132	-0.0101	-0.0147	-0.0194	-0.0268
	(0.0115)	(0.0112)	(0.0136)	(0.0130)	(0.0246)	(0.0247)	(0.0172)	(0.0167)	(0.0186)	(0.0180)
N	18981	18981	12442	12442	5618	5618	9062	9062	7460	7460
Panel B: Origin + destination fixed effects										
University enrollment	-0.0225	-0.0236	-0.0250	-0.0259	-0.0152	-0.0171	-0.0287	-0.0315	-0.0281	-0.0313
	(0.0064)	(0.0063)	(0.0067)	(0.0065)	(0.0194)	(0.0194)	(0.0092)	(0.0090)	(0.0082)	(0.0080)
Secondary school diploma in 5 years	-0.0222	-0.0224	-0.0253	-0.0251	-0.0086	-0.0096	-0.0168	-0.0189	-0.0366	-0.0388
	(0.0066)	(0.0061)	(0.0067)	(0.0062)	(0.0196)	(0.0189)	(0.0096)	(0.0092)	(0.0084)	(0.0080)
Years of schooling	-0.0272	-0.0277	-0.0305	-0.0309	-0.0102	-0.0109	-0.0266	-0.0298	-0.0306	-0.0341
	(0.0062)	(0.0059)	(0.0064)	(0.0061)	(0.0197)	(0.0194)	(0.0083)	(0.0082)	(0.0079)	(0.0076)
N	31333	31333	24777	24777	6522	6522	15921	15921	15469	15469
Cohort fixed effects	x	x	x	x	x	x	x	x	x	x
Individual characteristics	x	x	x	x	x	x	x	x	x	x
Age at move fixed effects	x	x	x	x	x	x	x	x	x	x
Only moved once					x	x	x	x	x	x
Times in difficulty before moving		x		x		x		x		x

Notes: Coefficients shown in the table are convergence rates β. Standard errors are clustered at the destination census tract level. Columns (1) and (2) include all students moving across census tracts within Montreal. In columns (3) and (4), the sample is restricted to census tract movers who moved across a FSA boundary at least once, and in columns (5) and (6) it is restricted to those who never moved across FSAs. Columns (7) and (8) restrict the sample to student who moved only once, and in columns (9) and (10) the sample is restricted to those who moved across FSAs exactly once (but may have moved across census tracts within FSAs multiple times).

Table E3: Decomposition: Census Tract Level Analysis

Sample:	All movers		
	(1)	(2)	(3)
University enrollment			
	Total exposure effects		
β	-0.0225	-0.0225	-0.0225
	(0.0064)	(0.0064)	(0.0064)
	School and non-school components		
β^{school}	-0.0158	-0.0161	-0.0118
	(0.0048)	(0.0048)	(0.0035)
$\beta^{non\text{-}school}$	-0.0067	-0.0064	-0.0107
	(0.0040)	(0.0041)	(0.0043)
Share school effects (S^{school})	70%	72%	53%
	(0.1394)	(0.1418)	(0.1041)
Secondary school diploma in 5 years			
	Total exposure effects		
β	-0.0222	-0.0222	-0.0222
	(0.0066)	(0.0066)	(0.0066)
	School and non-school components		
β^{school}	-0.0149	-0.0140	-0.0168
	(0.0052)	(0.0052)	(0.0062)
$\beta^{non\text{-}school}$	-0.0073	-0.0083	-0.0054
	(0.0038)	(0.0038)	(0.0039)
Share school effects (S^{school})	67%	63%	76%
	(0.1348)	(0.1354)	(0.1633)
Years of education			
	Total exposure effects		
β	-0.0272	-0.0272	-0.0272
	(0.0061)	(0.0061)	(0.0061)
	School and non-school components		
β^{school}	-0.0233	-0.0230	-0.0200
	(0.0052)	(0.0051)	(0.0044)
$\beta^{non\text{-}school}$	-0.0039	-0.0041	-0.0072
	(0.0033)	(0.0034)	(0.0035)
Share school effects (S^{school})	86%	85%	73%
	(0.1087)	(0.1120)	(0.0971)
Measure of school quality	$\pi\Omega_{s(n)}$	$\pi\Omega^{-i}_{s(n)}$	$\pi\Omega^{-i}_{s(n)}$
π	1	1	RD estimate

Notes: Sample restricted to movers across census tracts within Montreal. Standard errors are clustered at the destination census tract level, and obtained by the delta method. Parameters β^{school}, $\beta^{non-school}$ and S^{school} are calculated using equations (12), (13) and (14). In columns (1) and (2), π is set to one. In column (3), π is given by the RD-IV estimates from the sample of census tract permanent residents.