AGE EFFECTS AND INFORMATION SHOCKS

AGE EFFECTS AND INFORMATION SHOCKS: A STUDY OF THE IMPACT OF EDUCATION POLICY ON STUDENT OUTCOMES

By

JUSTIN SMITH, B.A. (HONS.), M.A.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

For the Degree

Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2008) McMaster University
(Department of Economics) Hamilton, Ontario

Title:	Age Effects and Information Shocks: A Study of The Impact of Education Policy on Student Outcomes
Author:	Justin Smith, B.A. (Honours) (McMaster University), M.A. (Simon Fraser University)
Supervisors:	Professor A. Abigail Payne Professor Thomas Crossley Professor Philip DeCicca
Number of Pages:	xii, 164

Abstract

This thesis studies the impacts of school entry policy and information revelation on student outcomes using a sample of students from the province of British Columbia (BC), Canada. The questions examined by the first two essays arise from a policy used by many industrialized countries, whereby students born within a 1-year time span all begin school at the same time. This policy creates large differences in age among students in the same class, which are thought to affect their academic performance along a number of dimensions. In the first essay, I contribute to the literature by establishing the persistence in test score differentials among students in the same class who differ in age. I show that in grade 4 older students outperform younger students by a large margin in numeracy, reading and writing, an effect that persists to a lesser magnitude until grade 10. The persistence is strongest for the writing skill, and it is also much stronger for girls than for boys. The strength of the test score differential in grade 10 suggests that the effects of age could have more lasting effects on cognitive and labour market outcomes.

In the second essay, I take a closer look at *how* age affects outcomes, by disentangling the entry age effect from the test age effect. Nearly all studies in this literature interpret age-related differences in student outcomes as the

result of entry age, but because students who enter later are also older at every stage in compulsory schooling, the entry age effect has not been separated from the test age effect. Using a set of students entering school at the time of BC's dual entry experiment, I show that test age is largely responsible for age-related differences in the probability of repeating grade 3, and entry age is largely responsible for age-related differences in grade 10 numeracy and reading scores. I show further that having an extra year of schooling reduces the likelihood that a student repeats grade 3, but has a negligible impact on grade 10 test scores. Both the entry age and test age effects are stronger for boys than they are for girls.

The final essay examines whether school choices change when parents are exposed to a new source of information on school quality. I model the effect of new information on choices using a simple expected utility framework and show that parents will use the new information to make different choices if they do not perceive it to be too noisy and if they have poor prior information on school quality. Furthermore, they make increasing use of the new information as more observations become available, since it becomes a more accurate predictor of true quality. Using the sudden release of BC's new standardized testing regime, I then study whether there is empirical support for the model. I show that the likelihood of switching out of a school increases when a school performs worse on the test, and that enrollment into kindergarten responds positively to increases in test scores. The response becomes stronger when more test score observations are available. Finally, I show variance in the response among parents living in less-educated neighbourhoods and among those who do not speak English at home, suggesting that prior information does play a role in the information use.

Acknowledgements

First and foremost, I would like to thank my advisor Abigail Payne for all her help and support over the past few years. I am thankful to her for being infinitely helpful during the entire course of my doctoral degree, for opening countless doors which have substantially improved my life, and for never turning me away when I showed up at her office door unannounced. If all advisors were like Abigail, the quality of academics would be much higher than it is today.

I would also like to highlight the support I've received from the other members of my thesis committee, Phil DeCicca and Tom Crossley. Phil and Tom have gone above and beyond what I would have ever expected from thesis committee members, and I'm truly lucky to have had their help.

Of course, I would have never been able to complete a thesis without the support of family. I thank my mom Cheryl Smith, my dad Allan Smith, my sister Jen Smith, and my nephews Fisher Obillos and Elliot Obillos for being understanding of the things that go hand in hand with graduate student life, like as weeks without phone calls, infrequent visits, and the occasional request for money. I thank the late Dianne Makuz and Theresa Smith, two of my biggest supporters and favourite people, who I'm sad to say will not get to see me graduate.

Thanks to my friends at McMaster, Logan McLeod, Kadir Atalay, Simo Goshev, and Claire De Oliveira for attending the practice seminars, the real seminars, and helping me work through the limitless problems that have come up over the past 4 years. Having this great cohort of classmates has made graduate school a great experience, and I don't think I could have done it without them.

Last, but definitely not least, I thank Lisa Roscoe for supporting me in the last and hardest year of the PhD program. Lisa is the sweetest and most thoughtful person I have ever met, and I'm lucky to have her in my life. I'm looking forward to my future with Lisa, and I can only hope to be as helpful to her as she has been to me.

Contents

List of Tables

List of Figures

Introduction

Though the economics of education is one of the youngest fields in the discipline, it has generated a considerable amount of research, particularly in the very recent past. As Hanushek and Welch (2006) note in the preface to volume 1 of the Handbook of the Economics of Education, this recent surge is partly being powered by an increase in the availability of large education data sets, which include both survey and administrative varieties. It is also being powered by an increased interest by governments, who are taking an active role in the determination of student outcomes. A recent example is the U.S. No Child Left Behind Act of 2001, which aims to improve schools by increasing accountability, public provision of indicators of school performance, and school choice.

While research in the economics of education spans a wide variety of topics and links to a number of other fields, two of the most popular research areas have been the effect of school entry policy on student outcomes, and what economists now simply refer to as "school choice," which encompasses the many effects of increasing choice among schools. This research has provided valuable and interesting insights into both fields, but has nevertheless left some questions unanswered while simultaneously raising new ones. The focus of this thesis is to provide empirical and theoretical answers to some of these questions: The first two essays are geared at providing answers to two of the remaining questions in the effects of school entry policy, while the third attempts to increase our understanding of school choice.

The entirety of this thesis relies on data collected by the Ministry of Education of British Columbia (BC). The data contain information on all kindergarten to grade 12 students in the province from 1990 through 2004. I make heavy use of scores from BC's annual standardized test, the Foundation Skills Assessment (FSA), which tests all grade 4, 7 and 10 students in the areas of numeracy, reading and writing starting in 1999. Because the data contain a large number of observations over a large number of years, the ability to track students across schools and time, and precise geographic identifiers, it represents the ideal information source to study all three topics.

The study of the effect of entry rules on student outcomes relies on the fact that most industrialized countries outline specific rules that govern when a child enters school for the first time. Typically, students born within a calendar year all begin school at the same time, but some variants to this rule are that students born between November of year t and November of year $t+1$ all begin at the same time, or as is the case in New Zealand, all students begin on the day they turn five. What all of these policies have in common is that they create a large age variance among students in the same school entry cohort. For example, if all students followed the rules to the letter, the age difference will be 1 year between the oldest and youngest child in any given class. Research has shown that this age difference can have a profound negative effect on the short and long term outcomes of the youngest students in the class, such as test scores, the probability of repeating a grade, diagnosis of learning disabilities, participation in clubs, *etc.*

In the first essay, the goal is to examine the persistence of the effect of these entry policies on student test scores in early elementary school through late high school. While researchers have uncovered a large, consistent set of results on the effects of these policies on student outcomes in early elementary school, the research on its persistence is comparatively sparse and much less consistent. For the most part, the cause of both of these problems is the unavailability of data on school outcomes late in high school. The few studies that are able to look more closely at persistence have typically pieced together data from two different sources, raising questions about the use of the results in studying persistence. The main contribution I make to this literature is to provide estimates of the effect of school entry policy on test scores across three grades in early elementary through secondary school from a consistent data source. Secondary contributions are methodological. Due to the structure of the data, I am able to use the Regression Discontinuity Design, a powerful method for identifying causal effects, which I then compare to the *status quo* method from the literature, Instrumental Variables.

I find that the youngest students perform substantially worse on tests in grade 4, which persists but shrinks in magnitude until grade 10. The decline in the effect is similar across numeracy and reading scores, whereas for writing scores there is much less shrinkage, suggesting that early advantages have a particularly lasting effect on this skill. Interestingly, the persistence across all skills is stronger for girls than it is for boys; the effect size for girls is about twice as large. From an examination of students with test scores in grades 4 *and* 7, the explanation for the shrinking test score gap between the oldest and youngest student is attributed to older students performing worse over time rather than younger students catching up. Finally, when comparing the Regression Discontinuity method to the Instrumental Variables method, I find that the Instrumental Variables estimates are similar but slightly lower, raising the possibility that it contains bias.

The second essay is an extension of the first, with the goal of answering a new question raised by this line of research. Essentially all of the existing research in this area has attributed the differential in test scores between the oldest and youngest students to differences in entry age. The problem with this interpretation is that the oldest students at entry are also the oldest students in the class at every stage in compulsory schooling, and because of this ambiguity it is not possible to attribute the entire effect to entry age. Disentangling the entry age effect from the test age effect is important because it helps parents and policy makers make more informed decisions, which could help them improve the outcomes of the youngest students. I provide a strong focus on gender in this essay, since boys and girls are thought to have certain biological differences at young ages that could cause them to be affected differently by entry age and test age. Highlighting gender differentials is also crucial for parents and policy makers to be fully informed.

By studying a group of students entering school at the time of BC's "dual entry" experiment, I find that the age-related advantage in the probability that a student repeats grade 3 is mostly related to increases in test age, whereas age-related advantages in test scores in grade 10 are mostly related to increases in entry age. Furthermore, as a by-product of the estimation method, I show that spending one extra year in kindergarten yields a benefit in terms of the probability of repeating grade 3, while it provides no benefit to test scores in grade 10. When separated by gender, I find that the test age effect and entry age effect are much stronger for boys. The results suggest a substantial benefit to delaying entry by one year, especially for male children. Such a student would be older at entry and in every grade, meaning higher test scores and a lower probability of repeating a grade. The results also suggest that taking gender into account when deciding whether to delay entry or provide compensating investments for young children may help to close existing gender gaps where boys are performing worse than girls along a variety of dimensions.

The final essay switches gears, and focuses on one particular aspect of school choice. In many jurisdictions in North America, schools operate under what economists have called the "local monopoly" system. Because school attendance is directly linked to residential address, parents generally do not have any choice over which school their child attends unless they switch addresses, giving rise to a school's monopoly power. This local monopoly system has been used in part as an explanation for why students in public schools underperform, since under such a system schools face no incentive to be efficient in producing education. Advocates of increased choice among schools argue that allowing parents to choose among a variety of schools would provide the proper incentive for them to produce better outcomes or risk parents leaving the school. This has been described by Hoxby (2002), for example, as a tide that lifts all boats.

In this essay, I examine to what extent one particular assumption behind the argument of school choice advocates matters for parent behaviour. For competition between schools to improve outcomes, parents must have enough information to be able to identify which schools are good and which schools are bad. Without being able to fully determine the quality of any given school, parents may mistakenly interpret some schools as good when in fact they are bad, a phenomenon called the lemons problem since Akerlof (1970). Under the assumption that parents are not fully informed when choosing among schools, I examine the their mobility response to new information on the quality of schools in BC.

The literature on the effects of new information on parent school choice is quite small. In terms of direct evidence, what is known is that in the U.S. parents who are given more information about the quality of schools in their choice set tend to choose better schools, and that this choice depends in part on their probability of actually being allowed to attend the school. More indirectly, a separate literature has shown that increases in school test scores lead to increases in the price of nearby houses, implying that parents are gravitating at least partly towards areas with good schools. The existing literature essentially examines what happens when existing information is repackaged into an easier format and presented to parents, which can be thought of as a saliency effect. One of the main contributions of the third essay, on the other hand, is to examine what effect an entirely new source of information has on parent choices. Another contribution is to provide a theoretical model to guide the thought process of what happens when parents are presented with such information.

To identify the effects of the new information, the main analysis compares the switching rate of parents out of schools before and after a new standardized testing system was introduced to the province. Because no standardized test was previously disclosed to the public at the school level, this represents an entirely new source of information. To address the possibility that switching costs play a large role in determining mobility, I also look directly at parent choices in kindergarten, when mobility is arguably less restricted by costs.

In the theoretical model, I show that parents should react to a new piece of information if it accurately identifies quality, and if parents had poor quality information to begin with. Furthermore, parents should make more use of new information over time since more observations from the same source will combine to provide a more accurate representation of school quality. In the empirical section where I test the results of the theoretical model, I find that the probability that a given student leaves their school is negatively related to that school's test score after the release of the information, indicating that parents are indeed using it. As more tests are revealed, I show that this negative relationship gets stronger, which is evidence that parents make increasing use of the new information. Finally, the response is weaker if the student lives in a neighbourhood with low education and if English is not spoken at home.

Though the evidence suggests that parents use the information, the response is quite muted, and one of the major concerns is that this is caused by moving costs. By examining the time period after BC removed school catchment area borders, and by looking directly at kindergarten choice, I show that the effect does not increase by much, suggesting that the muted effects are the result of parents not using the information. This could be the result of either strong prior information, or a belief by parents that the test scores are too noisy to be useful in choosing a school.

References

Akerlof, George A (1970) 'The market for 'lemons': Quality uncertainty and the market mechanism.' *The Quarterly Journal of Economics* 84(3), 488–500

Hanushek, Eric, and Finis Welch (2006) 'Preface.' In *Handbook of The Economics of Education,* ed. Eric Hanushek and Finis Welch, vol. 1 of *Handbooks in Economics* (North-Holland)

Hoxby, Caroline M. (2002) 'School choice and school productivity (or could school choice be a tide that lifts all boats?).' NBER Working Papers 8873, National Bureau of Economic Research, Inc, April

Chapter 1

Can Regression Discontinuity Help Answer An *Age*-Old Question In Education? The Effect Of Age On Elementary And Secondary School Achievement

1.1 Introduction

In most education systems all students start school at the beginning of the school year. The first year of schooling for a child usually depends on the relationship between the student's birth date and a qualifying date. For example, it is common for a school year to start in September and for a student to start school if she will turn five by December 31. While such a rule is easy to administer, it potentially results in a one year maturity difference across the students entering school. This rule may result in the younger students having more difficulties in school than their older classmates. Previous research confirms that at early ages older students perform significantly better than younger students on such outcomes as grade repetition, test scores, and special education diagnoses.

The research is less definitive on the question of whether this age effect persists. Understanding the persistence of this effect is important for understanding longer term outcomes such as wages and employment.[1] Furthermore, if the effect lasts until the end of compulsory schooling, it will bear directly on the use of quarter of birth as an instrumental variable for educational attainment in estimators of the returns to education, like in Angrist and Krueger (1991). A relationship between date of birth and performance on test scores at the end of high school implies that educational attainment is not the only source for the relationship between quarter of birth and wages, raising doubts that quarter of birth qualifies as an exogenous instrument.

Much of the existing literature on this topic focuses on outcomes in elementary school. Bedard and Dhuey (2006), Datar (2006), Elder and Lubotsky (2007), Puhani and Weber (2006), and McEwan and Shapiro (forthcoming) all find an older-student advantage of 0.4 standard deviations (σ) or higher below grade 4. The handful of studies that examine persistence (Bedard and Dhuey, 2006; Elder and Lubotsky, 2007; McEwan and Shapiro, forthcoming) all present evidence that older students outperform younger students into grade 8 at a magnitude of about 0.2 standard deviations (σ). While this is evidence in favour of a persistent age effect, none can say for certain if it

[1] Dobkin and Ferreira (2007) have shown that date of birth is not related to wages in the US census. These reduced-form effects, however, do no isolate the impact of having started school late on wages since date of birth can affect many variables affecting wages.

vanishes by the end of high school because none have data on test scores into higher grades.[2] Furthermore, among the three persistence studies, there is a discrepancy in how the age effect evolves over time. While both Bedard and Dhuey (2006) and Elder and Lubotsky (2007) find an age effect that diminishes into the later grades, McEwan and Shapiro (forthcoming) find that it could increase, though their estimates for grade 8 are not robust to specification changes, probably due to small sample size.[3]

In this paper I provide robust estimates of persistence well into high school using a unique administrative dataset from the province of British Columbia (BC), Canada. The data contain many features that are not available in datasets used in the literature. First, the BC data contains test scores for grades 4, 7, and 10, permitting an examination of persistence from elementary through the middle of high school, which other studies have not been able to do. Second, because there are several adjacent cohorts of children writing the same test, I use the Regression Discontinuity (RD) design, which is arguably the most reliable way to identify the age effect. Third, because I have multiple test scores for a subsample of students, I can answer a question that has not received any attention in the literature: whether attenuation in the age effect in upper grades is due to younger students catching-up or older students

[2]Fredriksson and Öckert (2006) have scores from grade 9, but do not look at persistence directly. Their findings for grade 9 are similar to the other three studies using data on grade 8.

[3]There are also a number of non-economic studies: Allen and Barnsley (1993); Cahan and Cohen (1989); Hauck and Finch Jr. (1993); Cameron and Wilson (1990); Dickinson and Larson (1963); Dietz and Wilson (1985); Sweetland and DeSimone (1987); Crosser (1991); Kinard and Reinhertz (1986); Langer et al. (1984); Stipek (2002). With the exception of Cahan and Cohen (1989), these studies are not concerned with causal effects, and generally ignore issues of endogeneity.

falling behind. Fourth, the data contain observations on kids in both public and private schools, allowing the opportunity to examine differences in the effect across different school structures. Finally, sample sizes are large enough that all effects are estimated very precisely. These added features allow the most robust estimates of persistence, and the only estimates into the middle of high school.

I show that the test scores of older children exceed those of younger children by 0.27 - 0.41 standard deviations (σ) in grade 4, declining to about 0.14 - 0.20σ in grade 10. The persistence is stronger for writing test scores than for reading and numeracy. The reason for the attenuation into grade 10 is driven by negative gains by older students and zero gains by younger students as they progress through the grades. Finally, there is a large degree of heterogeneity in this effect across personal characteristics; in particular, persistence is stronger for girls and students attending private schools. The persistence of the age effect into the 10th grade strengthens the case that the age effect may ultimately translate into labour market differentials of adults as they exit compulsory school. It also strengthens the evidence against the use of quarter of birth as an instrument for education in wage regressions, as used by Angrist and Krueger (1991). Given that date of birth affects educational achievement, if achievement then has an effect on wages, estimates on returns to education using quarter of birth is endogenous.

Policy response hinges crucially on *how* age affects educational outcomes. The literature is currently debating two separate questions that try to uncover these mechanisms. The first debate relates to whether the older-student advantage results from a chronological age difference (the absolute age effect) or from a relative age difference (the relative age effect). While I cannot provide any answers to this question, Elder and Lubotsky (2007), Fredriksson and Öckert (2006), Cascio and Schanzenbach (2007) and Kawaguchi (2006) have all attempted to separate these effects, and combined they seem to indicate a much stronger absolute age effect. The second debate relates to whether differences in age at test time (the age-at-test effect) or differences in age at school entry (the age-at-entry effect) are responsible this effect. It is thought that age-at-test effects will diminish over time because maturity differences between older and younger students disappear into the older grades. Age-at-entry effects, however, could grow if a mechanism such as ability tracking propagates them forward. The shrinking age effect into high school suggests that maturity is converging between older and younger students, and that what the literature observes is a strong age-at-test effect.

1.2 Empirical Model

Estimating the effect of age on test scores is difficult because parents have some control over their child's age in the classroom, principally through delayed entry. Because the motivation to delay entry is related to unobserved factors that affect the child's outcomes, OLS will not identify age effects. To circumvent this problem, this study uses the Regression Discontinuity (RD) design, which takes advantage of exogenous rules that assign children to different ages based on the school entry cutoff date. In BC, students enter school in the calendar year in which they turn five, meaning that a child born on December 31 will be among the youngest in the class, while students born the next day will be among the oldest. Assuming date of birth is random in the immediate vicinity of the school entry cutoffs, a comparison of the performance of kids born just to the left and right of January 1 will identify the causal effect of age on test scores.

Figure 1.1 illustrates the identification strategy using data from several birth cohorts of grade 4, 7, and 10 students aggregated to day of birth cell means.[4] On the right are plots of age against day of birth, which show a smooth downward relationship except for big discrete increases in age across each of the January 1 cutoff dates, which occur because of the school entry laws. An identical relationship exists for standardized numeracy test scores, which are plotted against day of birth on the left. The presumption is that the discrete jumps in age cause the discrete jumps in test scores because birthdate assigns students randomly to each side of the discontinuity. Because of the random assignment, there should be no differences between these children aside from their age.

In light of the random sorting to each side of the discontinuity, simple comparisons of the performance of the average child on the left and right of the cutoff would be appropriate if children complied perfectly with the entry rules. In practice, however, compliance is not perfect because parents sometimes delay entry. The slight upslope in age of kids born towards the end of the year in Figure 1.1 illustrates that this is most common among kids born

[4]The graph for grade 10 contains only 2 entry cohorts because they were the only group not affected a policy change occurring around that time.

in December. This situation calls for the "fuzzy" RD design, which corrects estimates of the test score difference by dividing by the actual age difference observed at the cutoffs.[5]

The causal effect of age (age_i) on test scores (T_i) is obtained by running two-stage least squares (2SLS) on the following model:

$$T_i = \beta_0 + \beta_1 age_i + f(bd_i) + v_{0i} \tag{1.1}$$

$$age_i = \gamma_0 + \sum_{j=1}^{4} \gamma_{1j} cut_{ji} + f(bd_i) + v_{1i} \tag{1.2}$$

The variables cut_{ji} a set of binary indicators equal to 1 if a student i is born to the right of cutoff point j, and their coefficients equal the size of the age gap at each discontinuity. The function $f(.)$ is a cubic spline that controls for the clear downward-sloping, non-linear relationship between age/test scores and birthdate (bd_i) away from the discontinuity points.[6] Regressions are estimated separately for each grade (4, 7 and 10) and each skill (numeracy, reading, and writing).

The coefficient β_1 is a Local Average Treatment Effect (LATE) for kids complying with the original school entry rules (*i.e.* those who do not delay

[5]For a complete description of RD theory, see Hahn et al. (2001), van der Klaauw (2002) or Lee (2005).

[6]The results below were insensitive to higher order spline functions. Note that the function $f(bd)$ need not be the same in each Equation. I focus on the case where they are the same to highlight the school entry cutoffs themselves as instruments generating exogenous variation. Because $f(bd_i)$ is an approximation to the true relationship between the left-side variables and bd_i, I use robust standard errors clustered on bd_i to account for the possibility that it is misspecified. See Card and Lee (2006) for details.

entry or repeat a grade). Normally the population to which this estimator applies is defined even more narrowly because it is only valid for those students near the cutoffs. I show later using a full-sample approach that this is a special case where RD identifies the effect even for those students not near the cutoffs.

1.3 Data and Summary Statistics

1.3.1 Data and Sample Description

Measures of student performance come from a full set of administrative records on the Foundation Skills Assessment (FSA) standardized test conducted by BC for students in grade 4, 7 and 10. These tests were initiated in school year 1999, and measure students' competency in reading, writing and numeracy. Scores are observed through the 2003 school year except the grade 10 test, which was discontinued for students entering 10th grade in that year.[7] Students are tracked over the years through a personal identification number, though if a student leaves BC or drops out, they are no longer observed. Students from all schools funded by the government, encompassing all public schools and some private schools have a test score.[8]

Test scores are linked to a large set of student-level administrative records that allow me to track a set of personal characteristics for every year back

[7]The British Columbia Ministry of Education changed their graduation program, which previously applied to grade 11 and 12 students, to include 10th graders. The graduation program has a different set of standardized tests, separate from the ones analyzed in this paper. The policy change does not affect the results presented here.

[8]Private schools are called "independent" schools in BC. I refer to them as private schools in this document.

to 1990. This gives complete histories for all grade 4 and 7 students, but slightly censored histories for grade 10 students. Not included in the personal characteristics is information on parental income and education. To proxy for these, I link Canada Census Profile data on income and education levels in small neighbourhoods to student records via their postal code.[9]

Though the data is collected by year and grade, the analytical sample is ordered into five consecutive *birth cohorts* for each of the grade levels 4 and 7, and two consecutive birth cohorts for grade 10.[10] The ordering of observations in this fashion is depicted in Figure 1.1. The data must be ordered in this way to implement the RD method, since calculating discontinuities involves comparing age and test score gaps between adjacent birth cohorts.[11]

For grade 4, the final sample is limited to those born in 1990 though 1994; for grade 7, it contains all students born in 1987 through 1991. Grade 10 uses two cohorts, which includes all students born in May 1986 through the

[9] A description of the procedure for this link is available from the author. Neighbourhoods are defined by a Dissemination Area (DA), a geographical area created by Statistics Canada containing a population between 400 and 700 people. DAs are the smallest level of aggregation for which data are distributed publicly, that still contain geographic identifiers. Their boundaries are defined naturally by things such as roads, railway tracks, waterways and power transmission lines. To put DAs into perspective, they are one level up in aggregation from a city block.

[10] There are only two cohorts for grade 10 because a policy change for the earliest three cohorts invalidates their use for this study.

[11] The birth cohorts are backed-out using the birth date for each student. This procedure works well for students who should have taken the test in year 2000, since even if they delay entry or repeat a grade, they will have a test score observation in 2001. For students who are supposed to take the test in 2003, however, any delayers or repeaters will have no test score observation because the data do not contain observations beyond 2003. Thus, the later birth cohorts will contain only those students who progress normally through school, as depicted by the perfectly straight line in age for the final birth cohorts in Figure 1.1.

Since delayed enrollment and grade repetition are most common among students born at the *end* of a calendar year, those very close to the discontinuity (at the beginning of the year) in the final cohort are still representative of the average student in other birth cohorts. In practice, estimates did not change with the exclusion of the final birth cohort.

end of 1987. For all grade levels, only students who have a positive test score attending a standard public or private schools are kept in the sample. Final baseline sample sizes are 225,127 in grade 4; 232,270 for grade 7; and 68,978 for grade 10. Furthermore, there is an auxiliary sample of 82,573 students who have a score in both grade 4 and 7.

Table 1.1 shows a set of means for students in each grade born in the first and last quarter of the year to highlight differences by date of birth relative to the January 1 cutoff. In Panel A, test scores have been standardized by year, skill, and grade, so that the units are standard deviations.[12] Across all grades students born in quarter 1 have higher test scores, which are large in the 4th grade but diminish into grade 10. There is some variance across skills, with numeracy and reading showing larger differences in grade 4, though this disappears by the 10th grade. Based on the these basic measures, it appears as though there is likely still a significant difference between older and younger children up until the 10th grade.

Panel B shows the differences in age expected by the school entry laws, and highlights the fact that the youngest students in the class delay entry and repeat grades more frequently which is the primary cause of bias in OLS regressions. Panels C and D show that other baseline personal characteristics are generally not different across birth quarters, lending support to the assertion that birth date is randomly assigned and the age effect is identified.

[12]Also note that the standardization was performed on the full sample, before the exclusions, which explains why the means for all students are not zero in Table 1.1.

1.4 Results

1.4.1 First-Stage and Reduced Form Estimates

Results from the first stage (FS) are in column 1 of Table 1.2, while the reduced form test score estimates are in columns 2-4. The predicted values from these regressions are plotted on top of cell means in Figure 1.2. The results for the first stage are from the sample of students taking the numeracy test, since they are essentially identical across all 3 skills.[13]

The first stage results estimate the age difference between the average student born on either side of each of the four cutoffs, and represent a measure of the degree of compliance with the school entry rules for students right at the cutoff. The average grade 4 student born just before the cutoff is about 0.78 years younger than the average person before the cutoff, reflecting the propensity of the December-born students to delay entry. The age gap shrinks in grade 7 and again in grade 10 because December-born students are more likely to be grade repeaters. The large F-statistics mean that the instruments are significant predictors of age.

The main result from the reduced form is persistent age gaps for all three skills. The reduced form ignores delayed entry and grade repetition, and measures the raw size of the test score gap at each cutoff. By grade 10 the estimates are about one-third the size of the 4th grade gaps, but still large. Quantitatively, within the 4th grade, the oldest students have about a 0.315σ advantage in numeracy, a 0.288σ in reading and a 0.219σ advantage in writing. This de-

[13]There are minor differences in the number of students taking the numeracy, reading and writing tests within each grade, so for brevity I have only reported those for numeracy.

creasing pattern from numeracy to writing is not as evident within grade 7, with the average discontinuity being 0.188σ for numeracy, 0.210σ for reading, and 0.193σ for writing. Within grade 10, all three skills show a difference of roughly 0.1σ. By comparison, Angrist and Lavy (1999) estimate that a class size reduction of 8 students leads to a 0.18σ - 0.29σ increase in test scores for students in the fifth grade. Furthermore, Rouse (1998) finds that students selected for the Milwaukee Parental Choice Program score about 0.08σ - 0.12σ better per year in math than similar students who did not take part in the program. Using these estimates as a benchmark, the age effect estimates reported here should be considered large.

At the grade 10 level, some of the youngest, least-ready kids have either delayed entry or repeated a grade and have therefore been re-sorted into the appropriate grade level for their ability for their age. It is therefore surprising that large reduced-form test score differences exist across birth dates at the cutoff. If these tools exist to make sure kids are ready enough to learn classroom material, then by the grade 10 level, kids should be re-sorted in such a way that the achievement gap between December and January kids is eliminated. Either delayed entry and repetition are not being used effectively, or this is the result of maximizing behaviour by parents who assume the cost of holding their child back by one year in terms of wages is higher than the potential benefits. Whether or not this behaviour actually makes the child better-off is not known, but it is likely the most plausible explanation for why the reduced-form test score gap still exists.

1.4.2 2SLS Estimates

The 2SLS results are presented in Table 1.3.[14] The general conclusion is a persistent age effect for all three skills. In grade 10, the numeracy and reading gaps are about one-third the size of the grade 4 gaps, while the writing gap is surprisingly stable at 60% of its size in grade 4. These results uncover the interesting fact that age does not affect all skills equally in terms of size or persistence. While the early impact of age differences on test scores fall disproportionately on quantitative skills, the persistence of the effect falls disproportionately on writing skills.

The 4th grade estimates are large at 0.407σ for numeracy, 0.375σ for reading, and 0.288σ for writing. While smaller, the grade 7 estimates are still large at about 0.260σ across all skills. The effects decline once again in magnitude by grade 10, but still remain large in magnitude and statistically significant at roughly 0.150σ for each skill. By comparison, the results from a simple OLS regression of test scores on age in Appendix Table 1.A1 show a strong *negative* effect in the upper grades.

The lower parts of each panel of Table 1.3 report results from two instrument tests. The first is the p-value from a J-test for the exogeneity of the instruments, which is possible in grades 4 and 7 because I am estimating a single age effect with information at four school entry cutoff points. The large p-values mean that the null of exogeneity is soundly rejected in almost every case, except grade 4 writing which seems to be an anomaly. The second is a

[14]Limited Information Maximum Likelihood (LIML) estimates were also computed, and are identical to the 2SLS estimates throughout.

weak instrument test developed in Stock and Yogo (2002). In their paper they list critical values for both the null is that 2SLS has large bias relative to OLS, and the null that tests based on 2SLS are of the incorrect size. The large value of the test-statistic implies in either case the null is soundly rejected and the instruments are strong.

It is important to know whether the age effect shrinks across grades levels because younger students gain on their older classmates or older students lose ground. Estimates at the bottom of Table 1.3 address this issue empirically by using test score *gains* as the dependent variable for a subsample of the grade 7 students who also have a grade 4 test score. The predicted values from these regressions are plotted in Figure 1.3 on top of cell means of test score gains at each date of birth. The structural model is the same as previous regressions, except the test score variable is replaced by the *difference* in test scores between grade 4 and 7. The results are uniformly negative across skills, meaning younger students have larger gains. Numeracy stands out as the largest and only statistically significant result at -0.169σ. Figure 1.3 plots the predicted values from the numeracy gains regressions on top of cell means by birthdate. At the discontinuity, the December-born students gain approximately zero, while older students have negative gains over time, meaning that the younger student advantage in gains is entirely due to older students losing ground.

To ensure age is the only variable affecting test scores at the discontinuity, the standard method in RD models is to check for discontinuities in baseline characteristics at the cutoff point. If baseline variables are smooth across the cutoff point, it provides strong evidence in favour of the random allocation of individuals to each side of the cutoff point. Tables 1.4 shows the results from a reduced-form RD model when the dependent variable is replaced by several baseline characteristics of students. The only statistically significant discontinuities are the grade 7 aboriginal status variable and the 10th grade ESL measure. In general, the number of statistically significant effects is no more than could be expected by random chance. Furthermore, the 2SLS estimates are relatively insensitive to the inclusion of the baseline variables in the regression, as seen by the even columns of Table 1.3.[15] Taken together, these results show no evidence that non-random sorting has affected the point estimates.[16]

1.4.3 Results by Subgroup

Table 1.5 calculates the 2SLS age effect by subgroups based on gender, ESL status, home language and public versus private schools.[17] The most striking result is that persistence is stronger for females than for males, in spite of the fact that grade 4 discontinuity estimates are similar between the two genders. The estimates for both males and females is roughly 0.400σ in grade 4 for numeracy and reading, and about 0.300σ for writing. In grade 10, however, the discontinuities for females are around 0.200σ across all skills, which is almost

[15]School fixed effects are included instead of the census measures since the census measures do not vary over time.

[16]An alternative test is to look for discontinuities in the distribution of births at the cutoff points like in McCrary (2007). Finding a mass of students born either before or after the cutoff would indicate sorting, and potentially invalidate results. I checked for discontinuities in the number of grade 1 students born at January 1 by running the RD model with number of students as the dependent variable. There was no evidence in favour of systematic sorting across the cutoff by parents. Results are available upon request.

[17]Only the estimates without covariates are discussed.

double the estimates for males. While the exact reason for this disparity is a puzzle, one possible explanation is even though the maturity gap between old and young females is the same as it is for old and young males at early ages, it is slower to converge for females as they get older.

Persistence in the age effect for ESL students is particularly low, even though it is similar to other groups in early grades. Once again, the reason for this is not obvious, but there are several competing explanations. First, the group of kids still taking ESL classes by the 10th grade might be a fairly homogeneous group of students who have a very hard time with the English language, so we would expect no differences in their school performance. Another possibility is that ESL students receive more one-on-one instruction, which mitigates any effect maturity might have had. The latter explanation is plausible given the result that students who are non-English speakers at home, an arguably similar group to ESL students, show a strong age effect into grade 10.

Finally, persistence is much larger for kids in private schools. This too could be the result of measuring gaps for a selected group of people, or it could be related to the differences in school structure between public and private schools. Perhaps parents pull high achieving students out of public schools and place them into private schools. If private schools give students better training, then these students will perform better relative to the others. Since most high-achieving students in early grades are January-born, private schools might contain a disproportionate number of very good students. On the other hand, maybe private schools are streaming kids while public schools are not.

Streaming kids tends to allow the best students to accelerate and the worst students to lag behind, and could help explain the widening of the age effect from grade 7 to grade 10. While I do not try to separate these hypotheses here, it would be interesting as a topic of future research to examine the structure of private schools and the propensity for parents to place their children into private schools depending on date of birth.

1.4.4 Full-Sample IV Results

Table 1.6 presents full-sample IV results from the model below to compare to the RD estimates,

$$T_i = \alpha_0 + \alpha_1 age_i + u_{0i} \tag{1.3}$$

$$age_i = \theta_0 + \theta age_i^p + u_{1i} \tag{1.4}$$

This method instruments actual age (age_i) with predicted age (age_i^p), equal to the age a student should be if she complied with the initial assignment at entry. The motivation behind this comparison is to check the external validity of RD away from the discontinuity point. A strong difference between the two estimators might imply that RD estimates apply only to students local to the discontinuity. The IV results are surprisingly similar to the RD results; the only difference is the IV estimates are slightly smaller in magnitude, but not by a large amount. This result makes sense if the youngest students benefit slightly more from being a year older at test time, which is likely to be true since they will also be the least mature students.

The difference in the estimates could also be driven by bias in the IV estimator, assuming that the RD estimates are unbiased and that age effects do not vary across birthdates. Bias in the IV method could be caused by birth date selection; the BC data show a strong seasonality of births, peaking towards the end of the summer. If a non-random group of parents selects summer births for their kids, then predicted age could be correlated with unobserved determinants of test scores, and the full-sample IV method would produce biased results.

To check for bias, I run reduced-form regressions of baseline covariates on predicted age; if birthdate is randomly assigned, the coefficient on predicted age should be zero. On the contrary, Table 1.7 shows that predicted age is significant in predicting several baseline characteristics, including aboriginal status, speaking a language other than English at home, and income - all factors that could conceivably impact test scores holding age constant. Even though the differences between RD and IV are not large, Table 1.7 suggests there is a potential for birthdate selection bias illustrating the potential superiority of RD

1.5 Interpretation

The literature is still unclear on exactly *how* age affects student outcomes, and there are two open questions currently being debated related to the mechanisms generating the age effect. The first is whether we are measuring an absolute age effect or a relative age effect. The absolute age effect arises from the simple chronological age difference between the oldest and youngest student. The relative age explanation states that older students perform better because they are the oldest in the class, regardless of how old students are when they take tests. In most studies these two effects cannot be separated because they are too collinear: kids who are absolutely older are also relatively older. Recently researchers have taken advantage of within-school, across-time variation in average age (Fredriksson and Öckert, 2006), cross-state variation in cutoffs (Elder and Lubotsky, 2007), random assignment of children to classrooms (Cascio and Schanzenbach, 2007), and exogenous changes in the number of births (Kawaguchi, 2006) to separate these two effects. While all of these studies estimate relative age effects that are fractions of the absolute age effect, none of them have been able to achieve statistical significance at conventional levels. Thus, their results must be taken with some caution.

The second debate relates to the contributions of age-at-test versus age-at-entry. Age-at-test arises because of age differences (whether relative or absolute) at the time they are examined, while age-at-entry is the cumulative effect of the age difference that existed at entry. These effects are also confounded because of a collinearity problem, since age-at-test is exactly equal to age-at-entry plus the number of years of schooling for everyone. While there has been no direct attempt to separate these two effects, they imply different evolutions of the age effect across time, so the persistence estimates are informative in which plays a larger role. Age-at-test effects are expected to converge over time because at each stage of education, maturity differences get smaller. As Elder and Lubotsky (2007) point out, they might also be converging because differences in pre-education human capital development matter less. Age-at-entry effects are expected to diverge over time because the older-student advantage is propagated by some mechanism, such as streaming or social promotion, making the initial advantage grow over time.[18]

While the results in the literature have been slanted towards an age-at-test explanation, McEwan and Shapiro (forthcoming) argue for age-at-entry effects because their estimates are potentially larger in grade 8 versus grade 4. Their grade 8 results, however, do not appear to be robust, since they are not consistent across math and science scores, and because they are very sensitive to the specification. Furthermore, using the same methodology I find a converging age effect that is robust across test subjects and specifications. Thus, given the bulk of the existing literature, and the robust estimates from this paper, I argue in favour of an age-at-test explanation.

1.6 Conclusion

This paper robustly estimates persistence in student test score differences caused by exogenous differences in age created by single-date compulsory school entry laws. The primary finding is that age effects last to a large degree well into middle-elementary and secondary school. This effect is not uniform across the distribution of personal characteristics, with females and private

[18]For a very good discussion of these mechanisms, see Bedard and Dhuey (2006). They give a good example of a situation where age at entry effects clearly dominate: Canadian hockey. This is because older players are larger, and are then streamed into more advanced levels from an early age.

school students showing particularly large persistence. The main implication is that school entry structure can have a lasting impact on student outcomes in the future. From a societal perspective it would be advantageous to correct the problem, since the systematic poorer performance of the youngest students represents foregone human capital acquisition, which could translate into negative labour market effects at the aggregate levels. Furthermore, the positive externalities generated by a more educated population would be increased.

There are competing explanations for the underlying cause, and it seems most likely age effects are the result of absolute age-at-test. Regardless of the underlying cause, the test score differences between the oldest and youngest student in the 10th grade are powerful evidence that quarter of birth is correlated with the unobserved determinants of wages. In reference to the famous article by Angrist and Krueger (1991), Bound et al. (1995) note even the slightest correlation between ability and quarter of birth will affect instrumental variables estimates when there is a weak correlation between the quarter of birth and educational attainment. The 10th grade estimates suggest that the correlation between ability and quarter of birth is actually very strong, and would likely affect coefficients in wage regressions even if the instruments were not weakly correlated with the endogenous variable.

What types of policies might correct the problem? One potential policy lever discussed by McEwan and Shapiro (forthcoming) and Elder and Lubotsky (2007) and Bedard and Dhuey (2007) is to move the cutoff dates earlier in the year. This policy has a differential effect on age in a class across the birth date distribution: it will make December-born kids absolutely and relatively older at entry, January kids relatively younger at entry, and will make the average student older. While it seems such a policy would benefit December-born kids, it could hurt the January-born, and therefore might not be an optimal policy. Whether this is true depends entirely on whether relative or absolute age effects drive the estimates from this paper. While they cannot estimate the effects across the birth date distribution, Bedard and Dhuey (2007) show that overall this policy has a positive effect on test scores and wages.

A second lever would be to keep the cutoff date the same, but increase both the entry age and exit age, so students start school later but stay in school for the same amount of time. While the relative age between the youngest and oldest kids would still be around 1 year, maturity differences between the oldest and youngest kids at any stage of school would likely be smaller. Furthermore, kids starting at an older age might be more "ready" to begin school. Both this and the previous strategy have the problem that they force students to postpone entry into the labour market, and if the costs of such a move are too high, then neither policy would be an improvement over the status quo.

References

Allen, Jeremiah, and Roger Barnsley (1993) 'Streams and tiers: The interaction of ability, maturity, and training in systems with age-dependent recursive selection.' *The Journal Of Human Resources* 28(3), 649–659

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., and Victor Lavy (1999) 'Using maimonides' rule to estimate the effect of class size on scholastic achievement.' *The Quarterly Journal of Economics*

Bedard, Kelly, and Elizabeth Dhuey (2006) 'The persistence of early childhood maturity: International evidence of long-run age effects.' *The Quarterly Journal of Economics* 121(4), 1437–1472

_ (2007) 'Is september better than january? the effect of minimum school entry age laws on adult earnings.' University of California Santa Barbara

Bound, John, David A. Jaeger, and Regina M. Baker (1995) 'Problems with instrumental variables estimation when correlation between the instruments and the endogenous explanatory variable is weak.' *Journal of the American Statistical Association* 90(430), 443–450

Cahan, S., and N. Cohen (1989) 'Age versus schooling effects on intelligence development.' *Child Development* 60(5), 1239 – 1249

Cameron, Mary Bridget, and Barry J. Wilson (1990) 'The effects of chronological age, gender, and delay of entry on academic achievement and retention: Implications for academic reshirting.' *Psychology in the Schools*

Card, David, and David S. Lee (2006) 'Regression discontinuity inference with specification error.' Number 322. In 'NBER Technical Working Paper.'

Cascio, Elizabeth, and Diane Whitmore Schanzenbach (2007) 'First in the class? age and the education production function.' Number 13661. In 'NBER Working Paper.'

Crosser, Sandra L. (1991) 'Summer birth date children: Kindergarten entrance age and academic achievement.' *Journal of Educational Research* 84(3), 140–146

Datar, Ashlesha (2006) 'Does delaying kindergarten entrance give children a head start?' *Economics of Education Review* 25(1), 43–62

Dickinson, Donald J., and J. Donald Larson (1963) 'The effects of chronological age in months on school achievement.' *The Journal of Educational Research*

Dietz, Carol, and Barry Wilson (1985) 'Beginning school age and academic achievement.' *Psychology in the Schools*

Dobkin, Carlos, and Fernando Ferreira (2007) 'Do school entry laws affect educational attainment and labor market outcomes?' University of California Santa Cruz and The Wharton School, University of Pennsylvania

Elder, Todd E., and Darren H. Lubotsky (2007) 'Kindergarten entrance age and children's achievement: Impacts of state policies, family background, and peers.' University of Illinois at Urbana-Champaign and Michigan State University

Fredriksson, Peter, and Björn Öckert (2006) 'Is early learning really more productive? the effect of school starting age on school and labor market performance.' Number 2006:12. In 'The Institute for Labour Market Policy Evaluation.'

Hahn, Jinyong, Petra Todd, and Wilbert van der Klaauw (2001) 'Identification and estimation of treatment effects with a regression-discontinuity design.' *Econometrica* 69(1), 201–09

Hauck, Anne Logan, and AJ. Finch Jr. (1993) 'The effect of relative age on achievement in middle school.' *Psychology in the Schools*

Kawaguchi, Daiji (2006) 'The effect of age at school entry on education and income.' Number 162. In 'ESRI Discussion Paper.'

Kinard, E. Milling, and Helen Reinhertz (1986) 'Birthdate effects on school performance and adjustments: A longitudinal study.' *Journal of Educational Research* 79(6), 366–372

Langer, Philip, John Michael Kalk, and Donald T. Searls (1984) 'Age of admission and trends in achievement: A comparison of blacks and caucasians.' *American Educational Research Journal* 21(1), 61–78

Lee, David S (2005) 'Randomized experiments from non-random selection in u.s. house elections.' University of California, Berkeley

McCrary, Justin (2007) 'Manipulation of the running variable in the regression discontinuity design: A density test.' NBER Technical Working Papers 0334, National Bureau of Economic Research, January

McEwan, Patrick, and Joseph Shapiro (forthcoming) 'The benefits of delayed primary school enrollment: Discontinuity estimates using exact birth dates.' *The Journal Of Human Resources*

Puhani, Patrick A., and Andrea M. Weber (2006) 'Does the early bird catch the worm? instrumental variable estimates of educational effects of age of school entry in germany.' Diskussionspapiere der Wirtschaftswissenschaftlichen Fakultät der Universität Hannover dp-332, Universität Hannover, Wirtschaftswissenschaftliche Fakultät, April

Rouse, Cecilia Elena (1998) 'Private school vouchers and student achievement: An evaluation of the milwaukee parental choice program.' *The Quarterly Journal of Economics*

Stipek, Deborah (2002) 'At what age should children enter kindergarten? a question for policy makers and parents.' vol. 16 of *Social Policy Report* (Society for Research in Child Development)

Stock, James H., and Motohiro Yogo (2002) 'Testing for weak instruments in linear iv regression.' NBER Technical Working Papers 0284, National Bureau of Economic Research, Inc, November

Sweetland, J., and P DeSimone (1987) 'Age of entry, sex, and academic achievement in elementary school children.' *Psychology in the Schools* 24(4), 406–412

van der Klaauw, Wilbert (2002) 'Estimating the effect of financial aid offers on college enrollment: A regression-discontinuity approach.' *International Economic Review* 43(4), 1249–1286

Table 1.1: Average Test Scores, Age, Personal and SES Characteristics by Birth Quarter

	Grade 4			Grade 7			Grade 10		
	All	Born Q1	Born Q4	All	Born Q1	Born Q4	All	Born Q1	Born Q4
Panel A: Test Scores									
Numeracy	0.004	0.130	-0.123	0.006	0.085	-0.072	0.021	0.052	-0.006
Reading	0.005	0.121	-0.116	0.006	0.087	-0.083	0.032	0.083	-0.001
Writing	0.004	0.095	-0.081	0.005	0.076	-0.066	0.030	0.069	-0.009
Panel B: Age									
Age	9.955	10.290	9.635	12.960	13.291	12.648	15.890	16.262	15.627
Repeated	0.023	0.017	0.038	0.069	0.058	0.097	0.113	0.069	0.135
Skipped	0.011	0.015	0.011	0.027	0.028	0.036	0.026	0.033	0.028
Delayed	0.017	0.002	0.057	0.018	0.004	0.058	0.016	0.002	0.043
Panel C: Personal Characteristics									
Male	0.505	0.508	0.503	0.504	0.506	0.501	0.498	0.501	0.496
Aboriginal	0.094	0.095	0.092	0.091	0.090	0.094	0.065	0.066	0.067
ESL	0.197	0.191	0.219	0.202	0.200	0.222	0.220	0.218	0.234
Non English	0.212	0.210	0.232	0.225	0.227	0.243	0.256	0.263	0.267
Private	0.149	0.149	0.151	0.154	0.153	0.157	0.156	0.160	0.155
Panel D: Census Characterisics									
Dropout	0.246	0.246	0.247	0.242	0.243	0.243	0.237	0.238	0.239
University	0.192	0.191	0.192	0.196	0.196	0.196	0.205	0.205	0.204
Income	62899	62894	62712	63724	63927	63530	64989	64904	64481

Test scores z-scores. Male = 1 if male; Aboriginal = 1 for Aboriginal; ESL =1 if in English as a second language program; Non-English =1 if home language is not English; Private = 1 if in private school; Repeated = 1 for grade repetition; Skipped = 1 for grade skipping; Delayed = 1 if delayed enrollment; Age = age in years; N = observations. Indicators note whether a student has *ever* had that trait. Dropout = % in neighbourhood with max level of schooling $<$ grade 9; University = % in neighbourhood with max level of schooling = some university. Income = avg household income of the neighbourhood.

Table 1.2: First Stage and Reduced Form Effect of Age on Test Scores

	First Stage	Reduced Form		
	Age	Numeracy	Reading	Writing
Grade 4				
Cut 1	0.749***	0.292***	0.264***	0.225***
	(0.008)	(0.027)	(0.027)	(0.029)
Cut 2	0.750***	0.316***	0.273***	0.153***
	(0.009)	(0.031)	(0.027)	(0.031)
Cut 3	0.816***	0.326***	0.339***	0.233***
	(0.007)	(0.031)	(0.029)	(0.028)
Cut 4	0.783***	0.329***	0.284***	0.270***
	(0.008)	(0.028)	(0.030)	(0.029)
Additional Stats				
Adj. R^2	0.645	0.010	0.008	0.005
F-Stat	24215.761	130.918	115.255	71.319
Grade 7				
Cut 1	0.739***	0.157***	0.173***	0.193***
	(0.010)	(0.026)	(0.024)	(0.027)
Cut 2	0.733***	0.213***	0.211***	0.189***
	(0.009)	(0.026)	(0.024)	(0.027)
Cut 3	0.792***	0.178***	0.201***	0.171***
	(0.009)	(0.026)	(0.027)	(0.023)
Cut 4	0.756***	0.200***	0.247***	0.214***
	(0.009)	(0.024)	(0.024)	(0.023)
Additional Stats				
Adj. R^2	0.588	0.004	0.004	0.003
F-Stat	19024.009	58.737	69.289	47.779
Grade 10				
Cut 1	0.717***	0.099**	0.111***	0.159***
	(0.012)	(0.031)	(0.030)	(0.028)
Additional Stats				
Adj. R^2	0.530	0.001	0.001	0.001
F-Stat	7870.534	7.672	14.924	16.766

* p<0.05, ** p<0.01, ***p<0.001. Test scores are z-scores. Standard errors are clustered on birthday. Models contain a cubic spline function of birthday. First stage estimate is from the numeracy sample; results using other skills are identical. F-stat tests significance of all regressors. "cut" variables refer to school entry cutoff points.

Table 1.3: 2SLS Effect of Age on Test Scores

	Numeracy		Reading		Writing	
Grade 4						
Age	0.407***	0.395***	0.375***	0.364***	0.288***	0.277***
	(0.021)	(0.018)	(0.020)	(0.018)	(0.021)	(0.018)
Tests						
J-Test (p-val)	0.897	0.966	0.539	0.327	0.081	0.032
Cragg-Donald						
F-Stat	19566	20110	19656	20225	19701	20260
Grade 7						
Age	0.246***	0.238***	0.276***	0.268***	0.254***	0.245***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Tests						
J-Test (p-val)	0.260	0.657	0.139	0.152	0.434	0.261
Cragg-Donald						
F-Stat	15679	16146	15646	16123	15716	16194
Grade 10						
Age	0.138**	0.106*	0.154***	0.150***	0.221***	0.202***
	(0.045)	(0.041)	(0.041)	(0.041)	(0.040)	(0.041)
Tests						
Cragg-Donald						
F-Stat	10647	10695	11240	11304	11188	11261
Grade 4 $\rightarrow$ *Grade 7*						
Age	-0.169***	-0.181***	-0.034	-0.037	-0.052	-0.044
	(0.026)	(0.029)	(0.025)	(0.030)	(0.036)	(0.039)
Tests						
Cragg-Donald						
F-Stat	35254	35415	34805	35017	34634	34764
Controls	no	yes	no	yes	no	yes

* p<0.05, ** p<0.01, ***p<0.001. Test scores were converted to z-scores prior to analysis, and are in standard deviation units. Robust standard errors clustered on birthday are in parentheses, except for the specifications with baseline controls, which are not clustered. All models contain a continuous cubic spline function of birthday, with a first stage regression that also contains a cubic spline function. Baseline controls included in even columns include *ever* male, Aboriginal, ESL, Non-English home language, attended private school and school fixed effects.

Table 1.4: Discontinuities in Baseline Characteristics at Enrollment Cutoffs

	Male	Aboriginal	ESL	Non English	Private	Dropout	University	Income
Grade 4								
Cut 1	0.005	-0.001	0.003	0.007	-0.001	0.002	-0.002	-956.220
	(0.013)	(0.008)	(0.011)	(0.012)	(0.009)	(0.003)	(0.003)	(722.443)
Cut 2	0.020	-0.000	-0.008	-0.004	-0.001	-0.002	0.003	463.851
	(0.013)	(0.009)	(0.011)	(0.011)	(0.010)	(0.003)	(0.003)	(725.282)
Cut 3	0.006	-0.001	-0.006	0.005	0.006	0.007	-0.002	127.804
	(0.012)	(0.008)	(0.012)	(0.012)	(0.010)	(0.004)	(0.004)	(686.571)
Cut 4	0.003	-0.005	-0.013	0.009	0.008	0.004	-0.001	-85.913
	(0.014)	(0.008)	(0.012)	(0.012)	(0.010)	(0.003)	(0.004)	(617.938)
Significance Tests								
F-Stat	0.630	0.120	0.459	0.324	0.216	1.474	0.564	0.609
Grade 7								
Cut 1	-0.013	0.000	-0.010	-0.000	-0.002	-0.004	0.003	766.301
	(0.013)	(0.008)	(0.012)	(0.013)	(0.009)	(0.003)	(0.003)	(642.805)
Cut 2	0.011	-0.016*	0.007	0.014	-0.006	-0.001	0.000	449.533
	(0.012)	(0.008)	(0.013)	(0.013)	(0.011)	(0.003)	(0.003)	(718.879)
Cut 3	-0.018	-0.000	0.007	0.015	-0.003	0.001	0.001	286.045
	(0.013)	(0.007)	(0.012)	(0.011)	(0.010)	(0.002)	(0.003)	(592.528)
Cut 4	0.009	0.003	0.010	0.012	0.002	0.001	0.000	67.290
	(0.013)	(0.007)	(0.012)	(0.013)	(0.009)	(0.003)	(0.003)	(757.571)
Significance Tests								
F-Stat	1.300	1.228	0.482	0.882	0.120	0.510	0.259	0.455
Grade 10								
Cut 1	-0.008	-0.007	0.025	0.046**	0.002	0.002	0.001	506.005
	(0.014)	(0.008)	(0.014)	(0.015)	(0.012)	(0.003)	(0.003)	(709.311)
Significance Tests								
F-Stat	0.301	0.624	3.336	9.599	0.032	0.498	0.048	0.509

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. See notes for Table 1.2 for regression details.

Table 1.5: 2SLS Effect of Age on Test Scores - by Subgroup

Statistic	Numeracy	Reading	Writing
Grade 4			
Male	0.416***	0.407***	0.281***
	(0.031)	(0.030)	(0.029)
Female	0.399***	0.349***	0.301***
	(0.027)	(0.026)	(0.026)
ESL	0.435***	0.344***	0.288***
	(0.043)	(0.041)	(0.039)
Home Lang Non-English	0.406***	0.390***	0.287***
	(0.023)	(0.023)	(0.024)
Public	0.409***	0.377***	0.274***
	(0.022)	(0.021)	(0.022)
Private	0.393***	0.345***	0.382***
	(0.055)	(0.056)	(0.065)
Grade 7			
Male	0.253***	0.247***	0.196***
	(0.029)	(0.030)	(0.027)
Female	0.240***	0.300***	0.301***
	(0.024)	(0.022)	(0.024)
ESL	0.249***	0.253***	0.253***
	(0.037)	(0.037)	(0.039)
Home Lang Non-English	0.239***	0.276***	0.244***
	(0.021)	(0.021)	(0.022)
Public	0.252***	0.274***	0.237***
	(0.020)	(0.020)	(0.019)
Private	0.212***	0.297***	0.375***
	(0.060)	(0.059)	(0.062)
Grade 10			
Male	0.086	0.105	0.187**
	(0.059)	(0.061)	(0.067)
Female	0.184**	0.190**	0.240***
	(0.065)	(0.061)	(0.060)
ESL	0.014	0.077	0.158
	(0.102)	(0.086)	(0.087)
Home Lang Non-English	0.140**	0.204***	0.234***
	(0.046)	(0.051)	(0.047)
Public	0.112*	0.111*	0.204***
	(0.044)	(0.045)	(0.042)
Private	0.300*	0.460**	0.326
	(0.144)	(0.147)	(0.167)

* $p<0.05$, ** $p<0.01$, *** $p<0.001$. See notes for Table 1.2 for regression details.

Table 1.6: Full Sample Instrumental Variables Effect of Age on Test Scores

	Numeracy		Reading		Writing	
Grade 4						
Age	0.387***	0.380***	0.363***	0.350***	0.274***	0.272***
	(0.009)	(0.008)	(0.009)	(0.008)	(0.009)	(0.008)
Tests						
Cragg-Donald						
F-Stat	402924	412926	405489	415690	405339	415459
Grade 7						
Age	0.247***	0.241***	0.264***	0.248***	0.222***	0.224***
	(0.009)	(0.008)	(0.009)	(0.008)	(0.009)	(0.008)
Tests						
Cragg-Donald						
F-Stat	327355	336713	327282	336697	328619	337913
Grade 10						
Age	0.089***	0.087***	0.127***	0.119***	0.115***	0.120***
	(0.016)	(0.014)	(0.016)	(0.016)	(0.016)	(0.014)
Tests						
Cragg-Donald						
F-Stat	83925	84089	87633	87836	87542	87799
Grade 4 → Grade 7						
Age	-0.158***	-0.149***	-0.073***	-0.067***	-0.061***	-0.051***
	(0.011)	(0.010)	(0.011)	(0.010)	(0.014)	(0.014)
Tests						
Cragg-Donald						
F-Stat	253820	255959	251828	254026	250142	251916
Controls	no	yes	no	yes	no	yes

* $p<0.05$, ** $p<0.01$, ***$p<0.001$. See notes for Table 1.2 for regression details. The instrument is predicted age, calculated by taking the age of a student if they had followed the entry rules. Baseline controls included in even columns include *ever* male, Aboriginal, ESL, Non-English home language, attended private school and school fixed effects. First stage is omitted.

Table 1.7: Effect of Predicted Age on Baseline Student Characteristics

	Male	Aboriginal	ESL	Non English	Private	Dropout	University	Income
Grade 4								
Predicted Age	0.003	-0.033***	-0.000	-0.024***	-0.004	-0.001	0.000	345.910
	(0.004)	(0.004)	(0.002)	(0.004)	(0.003)	(0.001)	(0.001)	(191.272)
	Male	Aboriginal	ESL	Non English	Private	Dropout	University	Income
Grade 7								
Predicted Age	0.007	-0.030***	-0.005*	-0.023***	-0.005	-0.001	-0.000	639.339***
	(0.004)	(0.003)	(0.002)	(0.004)	(0.003)	(0.001)	(0.001)	(192.439)
	Male	Aboriginal	ESL	Non English	Private	Dropout	University	Income
Grade 10								
Predicted Age	0.013	-0.029***	-0.002	-0.018*	0.001	-0.002	0.001	990.274**
	(0.007)	(0.007)	(0.003)	(0.007)	(0.005)	(0.002)	(0.002)	(358.647)

* p<0.05, ** p<0.01, ***p<0.001. Robust standard errors clustered on birthday are in parentheses. Predicted age is equal to the age a child should be if they comply with the school entry rules. See notes for Tables 1.1 for details on the variables listed in the column headings.

Figure 1.1: Average Age and Numeracy Test Scores by Birthdate

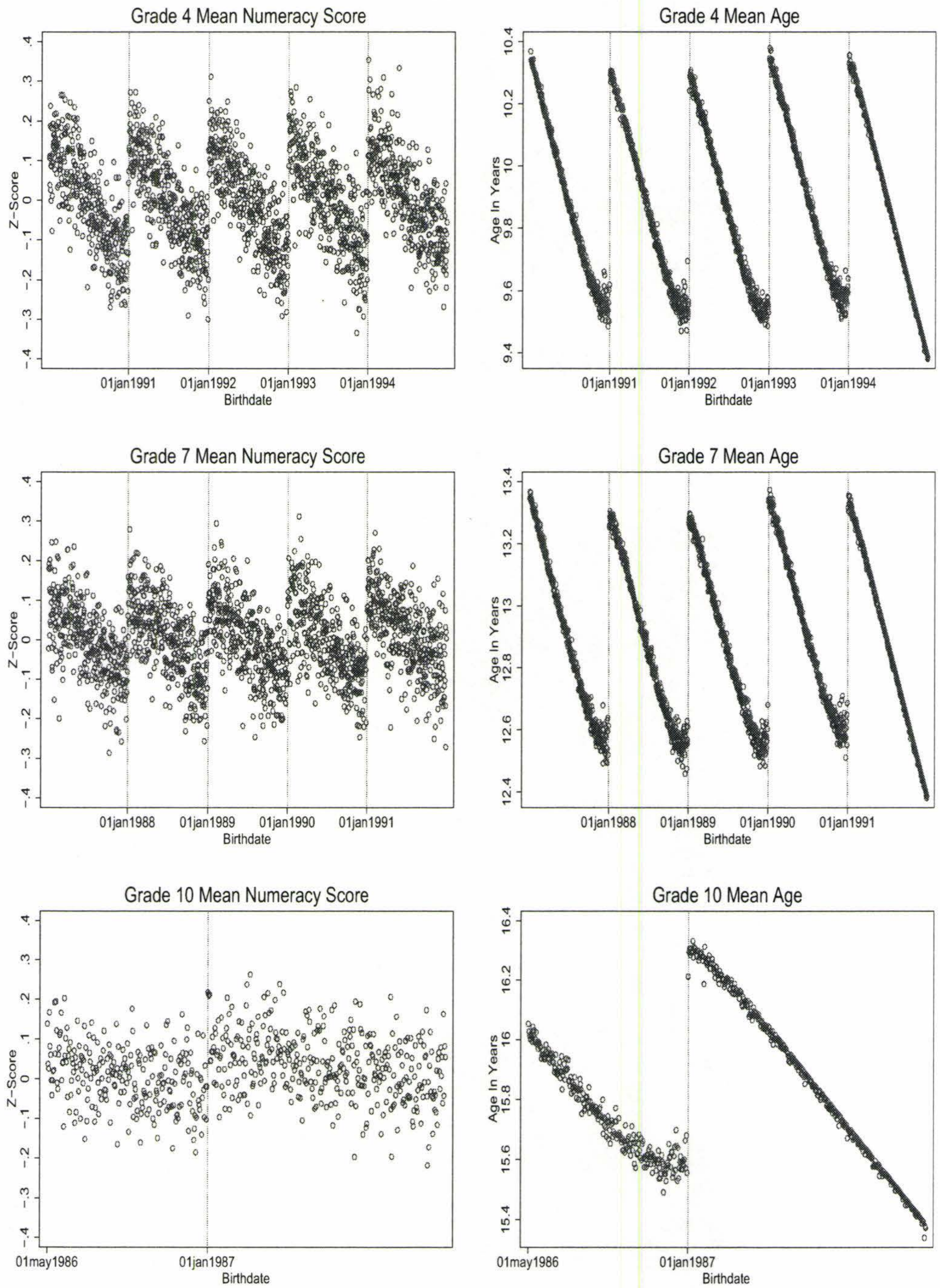


Figure 1.2: First Stage and Reduced Form Numeracy Estimates

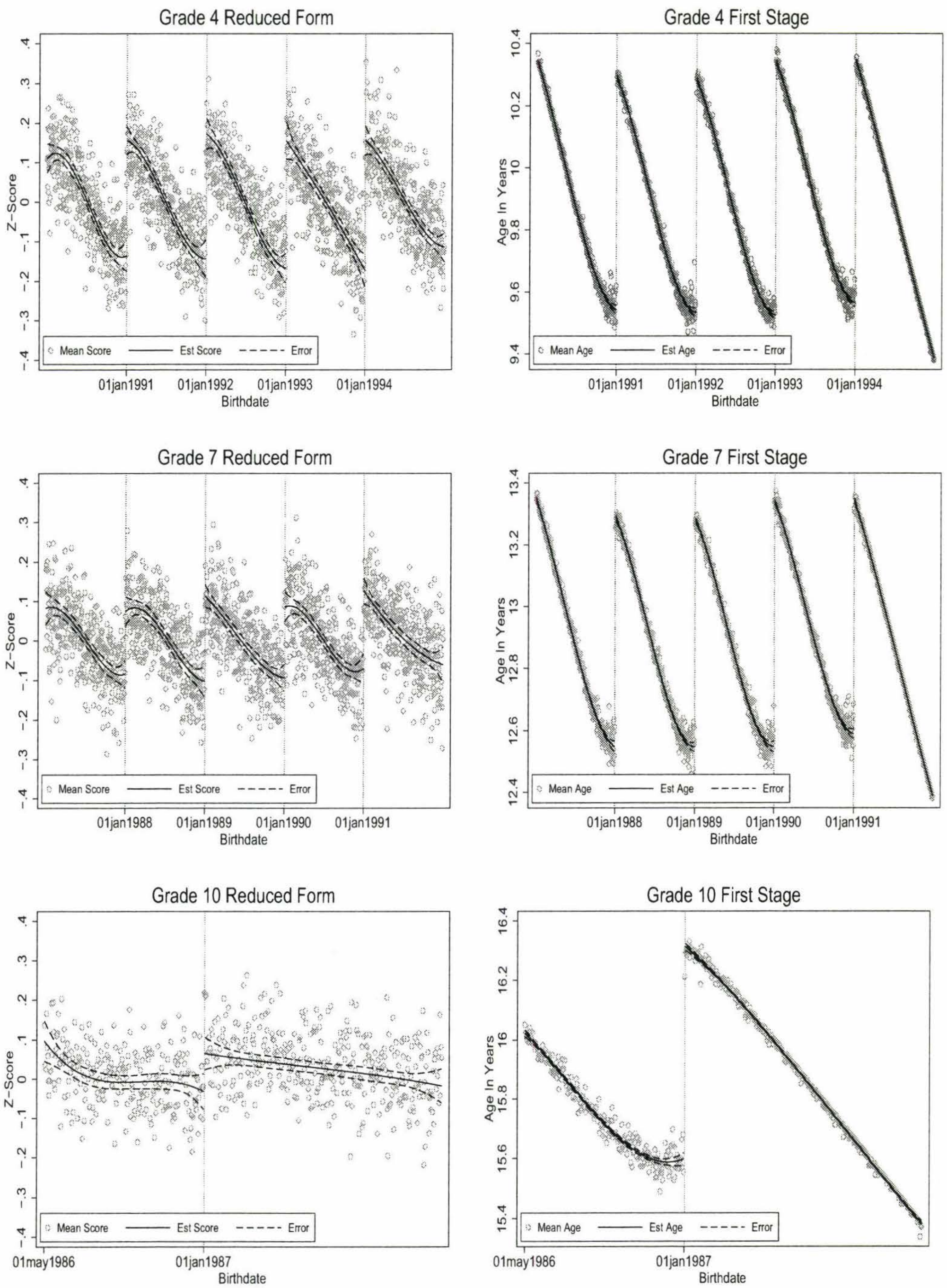
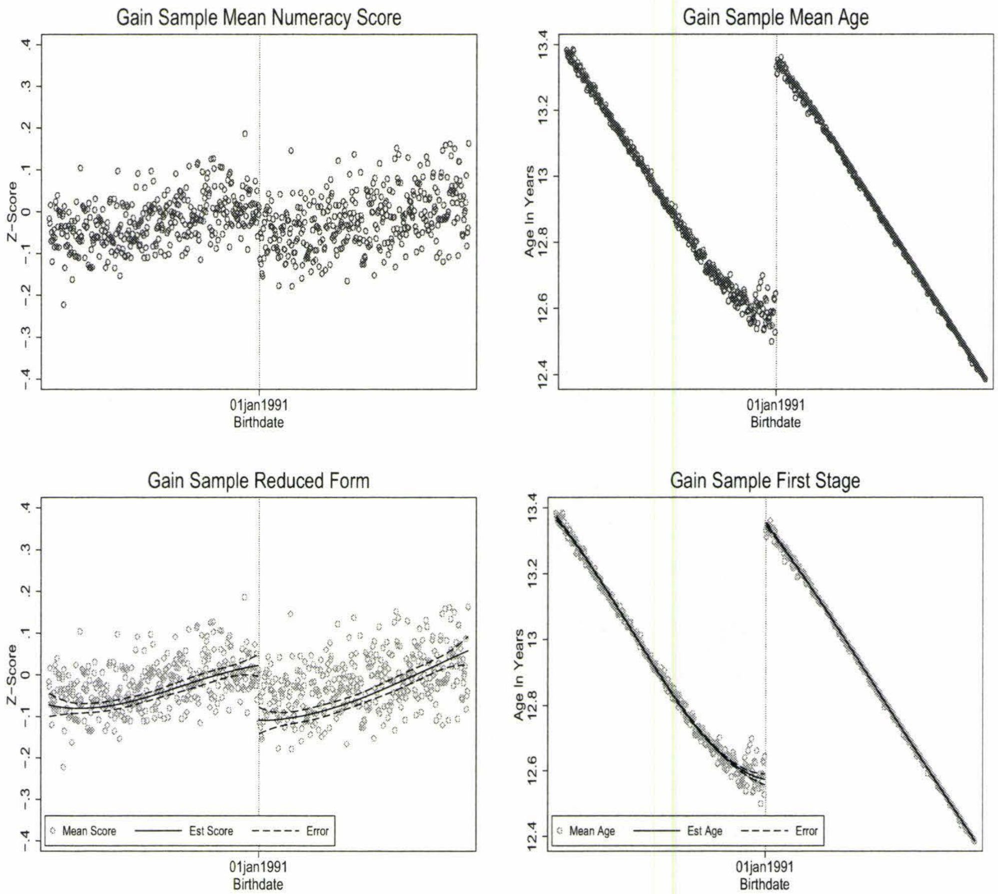


Figure 1.3: First Stage and Reduced Form Numeracy Gain Estimates



1.A Appendix: OLS Effect of Age on Test Scores

Table 1.A1: OLS Effect of Age on Test Scores

	(1)	(2)	(3)	(4)
Grade 4				
Numeracy	0.151***	0.155***	0.169***	0.169***
	(0.008)	(0.008)	(0.007)	(0.007)
Reading	0.119***	0.133***	0.129***	0.145***
	(0.008)	(0.008)	(0.007)	(0.007)
Writing	0.087***	0.121***	0.105***	0.132***
	(0.007)	(0.007)	(0.007)	(0.007)
Grade 7				
Numeracy	-0.014	-0.005	0.016*	0.016*
	(0.008)	(0.007)	(0.008)	(0.007)
Reading	-0.014	0.002	-0.001	0.017*
	(0.007)	(0.007)	(0.007)	(0.007)
Writing	-0.020**	0.016*	0.006	0.036***
	(0.007)	(0.007)	(0.007)	(0.007)
Grade 10				
Numeracy	-0.101***	-0.103***	-0.091***	-0.098***
	(0.014)	(0.013)	(0.013)	(0.013)
Reading	-0.127***	-0.107***	-0.129***	-0.105***
	(0.015)	(0.015)	(0.015)	(0.015)
Writing	-0.129***	-0.098***	-0.119***	-0.089***
	(0.015)	(0.014)	(0.014)	(0.013)
Baseline Variables	no	yes	no	yes
School Fixed Effects	no	no	yes	yes

* p<0.05, ** p<0.01, ***p<0.001. Standard errors are clustered on birthdate. Baseline controls included in even columns include *ever* male, Aboriginal, ESL, Non-English home language, attended private school and school fixed effects.

1.B Appendix: Attrition

A concern with the calculation of the test score discontinuities is attrition of under-performing students born in the last part of the year. Though school is compulsory for all students until the age of 16 in BC, there is still the possibility that students leave the school system and move elsewhere. However, these students may also be replaced by comparable students from outside BC who enter the school system. Appendix Table 1.B1 attempts to obtain more information about those students and whether their absence or entry is likely to alter the estimates of the age gap.

This exercise relies on the full set of K-12 observations. Appendix Table 1.B1 presents the percent of students in a grade who are ever observed in the future, broken down by birth quarter. If they are no longer observed, they have attrited. It also considers the reverse exercise, where students in a grade are checked to see if they are ever previously observed. If they are not, then they are new students. This set of results should indicate whether students leave school differentially by birth quarter, and if they are replaced by new students in the same proportions.

A larger proportion of the attrition in the early grades comes from the 4th quarter students, though this pattern reverses by the end of high school because these students reach the legal compulsory education floor. Students born in the first quarter attrit to a larger extent, which is reminiscent of Angrist and Krueger (1991), who use variation in schooling caused by this exact phenomenon. Unlike attrition, a higher proportion of new entrants always comes from the 4th quarter, especially in the 1st grade. This reflects late entrants who are more concentrated among students with 4th quarter birthdates.

Though middle-elementary 1st and 4th quarter attriters appear to be replaced in equal proportion by 1st and 4th quarter new entrants, there is a net loss of 1st quarter students and a net gain of 4th quarter students by the end of high school. Assuming the extra 4th quarter students are representative of the average 4th quarter student, and that the 1st quarter attriters are the lower performing students, there may be an abnormally high set of 1st quarter 10th grade test scores. It is reasonable to assume that the 1st quarter students who drop out are lower-performing, since as they reach the legal compulsory education age floor, those who have the least interest and least incentive to stay will be the first to drop out.

Researchers studying age effects should be cautious of this type of attrition, because it is likely not an idiosyncrasy in the BC data. For this study, attrition is not likely to be a problem in the 4th and 7th grades because attriters are being replaced in equal proportions in each birth quarter by new students. There may, however, be bias in grade 10.

Table 1.B1: Attrition and New Entry in BC Data

	% Attrited	Born Q1	Born Q4	% New	Born Q1	Born Q4
K	0.021	0.242	0.245			
Grade 1	0.019	0.239	0.249	0.062	0.225	0.315
Grade 2	0.018	0.242	0.248	0.036	0.241	0.249
Grade 3	0.019	0.237	0.248	0.035	0.242	0.253
Grade 4	0.018	0.236	0.249	0.035	0.244	0.245
Grade 5	0.017	0.241	0.245	0.030	0.247	0.252
Grade 6	0.017	0.236	0.245	0.032	0.249	0.249
Grade 7	0.019	0.246	0.248	0.035	0.245	0.252
Grade 8	0.021	0.246	0.243	0.034	0.237	0.255
Grade 9	0.027	0.253	0.245	0.038	0.245	0.256
Grade 10	0.044	0.253	0.239	0.051	0.246	0.257
Grade 11	0.073	0.254	0.238	0.054	0.242	0.258
Grade 12				0.029	0.244	0.250

Attrited refers to those students never observed in the data in any year after grade x. New refers to student who have never been observed in any year before grade x. In the calculation of the attrition (new student) measures, the last (first) year of data is dropped.

Chapter 2

How Valuable Is The Gift of Time? Disentangling Entry Age and Test Age In The Birth Date Effect

2.1 Introduction

A developing trend in North American elementary education today is what has been coined the "graying of kindergarten" - the increase in the age of the average student on their first day of school. For example, in 1968 96% of U.S. six year olds were in grade 1 versus 84% in 2005, with much of the difference being enrolled in kindergarten (Deming and Dynarski, forthcoming). Many jurisdictions legislate that children enter school in the calendar year they turn five, but an increasing fraction of parents, especially those with children that are young at entry, are exercising their option to delay entry and enroll their children in the year they turn six. The desire for children to be older at entry is based on the notion that older students perform better in school.

Indeed, this has been largely supported by the evidence, which shows that older students outperform younger students by a considerable amount along a variety of dimensions in early elementary school, persisting into late high school and into the labour market. Some U.S. states have taken note of this evidence, and used it as a reason to legislate school entry dates earlier in the year. Moving school cutoff dates earlier increases the average entry age, which should increase standardized test scores and give states a slight advantage in an outcomes-based accountability era.

What is almost always overlooked, however, is the fact that the oldest students at entry are also the oldest when tested in every grade, and we do not know whether the entry age effect or the test age effect is responsible for the advantages held by older students. Knowing the relative magnitudes of entering older versus being older when tested is important because it has implications for school policy. Finding a large difference in student performance attributable to entry age might mean rethinking systems of school entry where children can be up to a year apart in age on their first day of school. Finding a large difference in performance attributable to test age suggests finding alternatives to the common practice of testing all students in the same grade at the same time, but does not necessarily have direct implications for school entry policy. If a school was instead interested in directing some resources towards younger students to help them catch up to older students, the timing and duration will depend on the magnitude of entry age versus test age.

The relative magnitudes of entry age and test age help to inform parental behaviour. Parents typically use either grade repetition or delayed entry to exert some influence over their child's age in the classroom. Repeating a grade would only be beneficial if cognitive advantages held by older students are due to their higher test age, since holding a child back makes them one year older when tested. If cognitive advantages are related to higher entry age, such a strategy would yield no increase in outcomes. Delayed entry, on the other hand, would be beneficial regardless of whether entry age or test age carries more weight, because such a strategy makes a child older at entry and in all subsequent grades. Nevertheless, separating the effects may still be of interest to parents who want to know why delaying entry benefits their child. If parents make compensating investments in younger children, such as extra books or tutoring, instead of trying to manipulate their child's age in the classroom, the timing and duration of such investments depend on the relative magnitudes of entry age and test age.

It is also important to recognize how the effects of entry age and test age vary by gender, as there is good reason to think that they would be affected differently. From a biological point of view, at young ages there can be differences in brain development between young boys and girls, and therefore they might differ in their ability to learn certain material if they enter school at the same age.[1] In support of this argument, the same evidence that shows older students outperform younger students has also revealed heterogeneity by gender. Quantifying heterogeneity across gender is crucial because fully-informed parent and school policy decisions will need to consider these gender

[1]For example Sax (2007) describes that parts of the male and female brain have different developmental trajectories, and that, for example, the development of the language center of the male brain can be up to a year and a half behind the female brain.

differences. Furthermore, a careful consideration of gender in school entry decisions by parents and policy makers may help to close the recently-documented widening gender gap in student outcomes that has seen boys underperforming along a variety of dimensions, including test scores, grade repetition, dropout, and college attendance.

This paper disentangles entry age from test age and studies their effects on the probability of repeating grade 3 and the numeracy and reading test scores in grade 10 of a group of children that were part of an education experiment in British Columbia (BC), Canada.[2] In 1990, BC mandated a "dual entry" system, whereby students entered school in either September or January depending on their birth date. The policy was repealed one year later, and students were reshuffled into grade levels in such a way that the effects of entry age and test age on outcomes can be statistically separated, while also providing the opportunity to estimate the effect of one extra year of kindergarten.

I show that large differences in the probability of repeating grade 3 are mostly attributable to test age, and, contrary to previous research on the issue, differences in grade 10 numeracy and reading test scores are mostly attributable to entry age. I show further that one extra year of kindergarten has a negative impact on the probability of grade repetition, and small but negative impact on test scores. By summing the separate estimates of test age and schooling together, I can also mimic the effect of grade repetition on

[2]As I discuss in detail below, the reason why I study the repetition behaviour of grade 3 students only is that most of the individuals from this group who actually repeat a grade do so in grade 3.

future test scores, which turns out to be close to zero. Estimating these effects separately by gender reveals that the magnitude of the test age effect on grade repetition for boys is about double that of girls, and the entry age effect on test scores is almost entirely driven by boys. The mimicked effect of grade repetition on test scores is slightly negative across both skills for females, and close to zero for boys.

The results suggest that the parental practice of delaying entry will carry multiple benefits that carry into the long term. Because it makes children older at entry, a positive test score advantage is gained for both reading and writing in grade 10, especially for males. Furthermore, by being one year older at the beginning of each grade, the same child is much less likely to repeat a grade, especially if that child is male. The practice of repeating a grade, on the other hand, does not appear to offer any cognitive benefits, and could be detrimental if one considers that such a student delays entry into the labour market by one year. Finally, because cognitive advantages are mostly driven by entry age, the results also suggest that any compensatory investment into younger children's education by the parent or by the school should take place early, and special attention should be paid to boys if reducing gender inequality is a policy goal.

2.2 Entry Age or Test Age? The Identification Problem

Suppose that for each child i in grade g the following relationship holds between some educational outcome (O_{ig}), entry age (AE_i), test age (AT_{ig}), and schooling (S_{ig}),

$$O_{ig} = \beta_0 + \beta_1 AE_i + \beta_2 AT_{ig} + \beta_3 S_{ig} + \epsilon_{ig} \tag{2.1}$$

Given a set of data containing information on each of these four variables for children across several grades, it is impossible to estimate this equation because they are linked by the identity

$$AT_{ig} = AE_i + S_{ig} \tag{2.2}$$

The age effects literature has focused on trying to identify β_1 by regressing O_{ig} on AE_i using instruments generated by student birth dates and school entry cutoff rules. They abstract from the effects of S_{ig} by running regressions separately by grade, which implicitly holds S_{ig} constant. This strategy will fail to identify β_1, however, because for all students in the same grade AE_i is an exact linear function of AT_{ig}, and what is actually being identified is

$$O_{ig} = \beta_0^* + (\beta_1 + \beta_2) AE_i + \epsilon_{ig} \tag{2.3}$$

where $\beta_0^* = \beta_0 + \beta_2 g$, and g is the constant number of years of schooling.

A visual depiction of the identification problem is plotted on the left side of Figure 2.2 for a series of birth cohorts when they are in grade 3 under a single-entry school system. This graph plots both theoretical entry age and test age against birth date assuming that no students skip, repeat, or delay entry. For each student, the difference between entry age and test age is a constant, making it impossible to include both variables in the same regression.

Given age variation generated from single-date school entry laws, there is no easy way to break the link between AT_{ig} and AE_i. One method would be to compare the outcomes of grade repeaters to normal students who entered school at the same time, as they would have different values of AT_{ig}, but this is likely to be a bad strategy given that the two groups of students differ in many unobserved ways. The only potentially exogenous way to break the direct link is to make use of some policy intervention that assigns students to different ages at entry and at test time that is unrelated to unobservables.

2.3 Existing Literature

Most of the papers in the age effects literature have estimated that younger students perform worse than older students for a variety of educational outcomes, a result that is robust to changes in country, dataset, and estimation method. In early elementary school, researchers find that older students have a test score advantage between 0.3 - 0.8 standard deviations (σ), are less likely to repeat a grade by up to 13 percentage points, and are less likely to be diagnosed with a learning disability by up to 3 percentage points. Towards the end of compulsory schooling and beyond, researchers show that the test score advantage held by older students persists at a lesser magnitude of roughly 0.1σ, and that other outcomes such as being university-bound and wages are also affected by age.[3] While these studies interpret their estimates as the effect of starting school later, the identification problem outlined above suggests that they cannot separate it from the effect of being one year older when tested.

A handful of papers have made attempts to separate these effects.[4] An early attempt is Datar (2006), who uses the ECLS-K sample to estimate the entry age effect for a group of children in kindergarten and grade 1. Using the fact that ECLS-K is a panel dataset, the test age effect is differenced out by using test score gains as the dependent variable, which identifies entry age if test age is linear over time. She finds a significant positive entry age effects. This paper, however, is limited by the fact that it cannot provide test age effects, and because the data only allow a focus on very young students. Furthermore, it is unclear whether test age effects are linear over time, which is crucial to the identification strategy, and it is unclear what role schooling plays in the estimates.

Black et al. (2008) estimate the effects of entry age and test age on IQ scores for a sample of Norwegian men who are 18 years old. Because the school entry cutoff structure is not identically related to the age at which the IQ test is

[3]See Chapter 1 of this thesis, Bedard and Dhuey (2006), Elder and Lubotsky (2007), McEwan and Shapiro (2008), Fredriksson and Öckert (2006), Datar (2006), Elder and Lubotsky (2007), Puhani and Weber (2006), and Dhuey and Lipscomb (2007)

[4]Researchers are also trying to identify the separate effects of relative age. See (Fredriksson and Öckert, 2006), (Elder and Lubotsky, 2007), (Cascio and Schanzenbach, 2007), and (Kawaguchi, 2006). The general consensus seems to be that the relative age effect is small and likely very close to zero.

that the fraction of males who delay entry is consistently higher than females. To the extent that delayed entry is mostly related to school readiness or maturity, given the common perception that males mature slower than females this result is expected.

The exception to the general pattern in Table 2.3 is that only half of the students born between November and December, 1985 comply with the rules, another 30% start early (in September 1990) and the remaining 20% delaying entry (to September 1991). Males show a higher propensity to delay entry than females, consistent with the rest of the figures in the Table. Though early entry is technically not allowed by law, parents likely petitioned to have their child start in September 1990, since this is the date they would have started under a single-entry system. The high fraction of late starters might be related to a large group of parents that did not want their children to begin school half way through the school year in January.

This low level of compliance for November - December children does not bias the results because the estimation method uses an instrumental variables strategy that applies only to compliers. It will, however, affect who is driving estimates if November - December compliers differ from compliers born near other discontinuities; *i.e.* it implies that the LATE I am estimating will diverge more from the ATE. Table 2.4 presents the fraction male, aboriginal, and Non-English speaking compliers born within 1 month of each cutoff point to see if they are different along these characteristics. The general finding is that the compliers born near the November cutoff are very similar along all metrics to the other cutoffs, so the difference in the compliance rate should not affect the results.

Figure 2.3 looks the variables of interest at the discontinuities to confirm that there is enough variation to identify parameters. On the top left is a plot of average entry age, test age, and years of schooling by birth date for a group of grade 3 students. The variation mirrors the theoretical pattern given by from Figure 2.2 for all three variables, which is expected given the high compliance with school entry rules. The effect of the lower compliance by the November-December 1985 children has manifested itself as increased fuzziness in entry age and test age near the cutoff points. Identification depends on average entry age being continuous across the January 1, 1986 cutoff and test age being continuous across the November 1, 1985 cutoff, which appears to be satisfied in this graph.

The top right graph in Figure 2.3 plots the fraction of grade 3 repeaters at each birth date. There is a clear increasing pattern in the probability of repetition towards the end of the year, reflecting the fact that these students are on average the youngest in the class. There are, however, no noticeable discontinuities at November 1 or May 1. The large discrete jumps at all of the January cutoffs and lack of such jumps at the other cutoffs is preliminary evidence of a large test age effect and a small entry age effect. The logic behind this conclusion is that at all of the January cutoffs there are jumps in test age, but at the others there are not, indicating that once variation in test age is removed, so is variation in probability of grade repetition.

The bottom graphs plot average grade 10 numeracy (left) and reading (right) z-scores against birth dates. The jumps at the cutoffs are slightly less visually obvious than they were using grade repetition as an outcome, but there are still noticeable jumps at the end of each calendar year. Unlike the grade repetition graph, in these graphs there are potentially breakpoints in the average performance of students born to the left and right of November 1 and May 1, where entry age varies. These results are suggestive of both an entry age and test age effect.

Figure 2.4 examines the variation in baseline variables across the discontinuity points to provide evidence in support of the identification of α. RD designs rely on random assignment of students to each side of a set of cutoffs, and given baseline characteristics a straightforward way to check if this condition is met is to examine the behaviour of observable baseline characteristics across the cutoffs. The identification strategy is supported if all previously determined variables are continuous across cutoffs, since this implies a lack of sorting across cutoffs. The results from Figure 2.4 show that fraction Aboriginal, male, and non-English speakers are visibly continuous across all cutoffs, supporting identification.

2.8 Results

2.8.1 Regression Discontinuity Estimates

This subsection presents the RD estimates of α, which are used as a stepping-stone to obtain estimates of β. Table 2.5 displays first-stage estimates to examine the variation in entry age and test age around the cutoffs. Table 2.6 then shows the estimates of the three elements of α for grade repetition, numeracy scores and reading scores, along with some model specification tests. When grade repetition is the outcome the coefficients are decimals, and when test scores are the outcome the coefficients are in terms of standard deviations (σ).[13] All models in this section are estimated by 2SLS, with standard errors clustered on bd_i.

Table 2.5 shows the first stage estimates of the discontinuities in test age and entry age for both the repetition and test score samples. Column 1 provides estimates of the differences in test age at each of the three discontinuities. The results show that students born just after cut_{J85i} and cut_{J87i} are about 0.795 to 0.856 years older than students born just before, whereas at cut_{J86i}, students born just after the cutoff are only about 0.650 years older. Column 2 looks at differences in entry age at the various cutoffs, and estimates that students born just after cutoff cut_{N85i} are about 0.296 years older at entry, while at cut_{M86i} the number is approximately twice that size. The collinearity between entry age and test age is demonstrated by the similarity of the estimates at cut_{J85i} and cut_{J87i}, where both variables jump together. Columns 3 and 4, which use the grade 10 sample, show similar estimates to columns 1 and 2. Results by gender are presented in Appendix Table 2.A1. Estimates are similar, except that gaps at the discontinuities are higher for females than for males, reflecting the fact that males are more likely to repeat grades and to delay entry. In both Tables, the model fit is strong, and the first stage F-statistics are extremely high indicating strong instruments.

The 2SLS estimates of α for all three outcomes are presented in Table 2.6.

[13]Test scores were standardized to have a zero mean and unit standard deviation for each year and skill.

Recall that α_1 is the joint effect of test age and entry age, α_2 is the joint effect of test age and schooling, and α_3 is the joint effect of entry age and schooling. The parameter α_1 is the quantity that existing papers have called the entry age effect. α_2 has an interesting interpretation as the effect of repeating a grade on outcomes, since repeating a grade involves increasing test age and schooling by one year. α_3 on its own does not have any particularly useful interpretation.

The estimates of α_1 show that students born just after the cutoff are about 7.8 percentage points less likely to repeat a grade, and perform about 0.082σ - 0.091σ better on standardized tests, which is comparable to what has been found by others in the literature. Estimates of α_2 show that test age and schooling combine to have a negative effect on grade repetition, but essentially a zero effect on test scores. This latter result is interesting, because the combination of increases in test age and more schooling mimic what would happen if a student repeated a grade, and thus the estimate of α_2 on test scores implies that repeating a grade yields no benefit for students in terms of test scores. α_3 is also imprecisely estimated in all three regressions, and it is close to zero for grade repetition supporting further that entry age does not affect this outcome. Gender differences in the estimates of α are presented in Appendix Table 2.A2. The magnitudes are stronger for males than for females, especially α_2 for grade repetition and α_3 for test scores.

Because the estimation strategy relies heavily on instrumental variables, Table 2.6 presents some statistics that test the quality of the cutoff dummies as instruments for test age and entry age. The under-id F-stat tests the null hypothesis that the matrix of coefficients in the IV model has full rank versus the alternative that it does not, the weak-id F-stat test the null that the instruments are weak versus the alternative that they are not. Both null hypotheses are soundly rejected, indicating strong instruments.[14]

To check the exogeneity of the instruments near the cutoffs, Table 2.7 presents formal estimates of the discontinuities in baseline characteristics depicted in Figure 2.4.[15] Finding a zero coefficient supports the hypothesis that students are randomly allocated around the cutoffs. Also note that in this Table, small F-statistics are favourable, as they indicate that the cutoff dummies are poor instruments for the baseline variables, as they should be.

The general pattern is small, near-zero estimates of the discontinuities at nearly all of the cutoffs. The exceptions to the patter are discontinuities in the fraction male and fraction non-English. The fraction male is almost certainly the result of sampling variation, since parents generally do not have the option of selecting the sex of their child, and therefore this variable could not logically be manipulated by parents. The discontinuity in non-English speakers may be a sign of *some* sorting, but hardly evidence of a systematic pattern by parents. The low F-statistic indicates that these cutoff dummies are poor instruments for each of these baseline variables, especially given that the Staiger-Stock rule of thumb is for the F-stat to be above 10 for instruments to be considered good. Thus, overall the lack of a general pattern of discontinuities in this Table

[14]See Stock and Yogo (2002) for cutoff values for this test.

[15]Another option would be to perform an overidentification test, since I am overidentified in some equations. This is not a good idea because *a priori* I know that two variables jump at each discontinuity, and therefore I expect that the model would fail such a test. Instead I rely on the standard practice in RD papers to check for exogeneity, which involves the procedure described in this paragraph, and the inclusion of covariates in the RD regression.

provides support for the exogeneity of the cutoff dummies as instruments for entry age and test age.

2.8.2 Structural Parameter Estimates

Table 2.8 presents the independent effects of entry age, test age, and schooling on grade repetition (Panel A), numeracy (Panel B), and reading scores (Panel C). Within each panel, the first row pools all of the data, while in the subsequent two rows the results are separated by gender. Specifications including and excluding a set of school fixed effects and baseline variables are presented side-by-side to highlight the robustness of the estimates to the inclusion of these variables. The coefficients in Panel A are in terms of percentage points, while those in Panels B and C are in terms of standard deviations. Because analytic standard errors are difficult to obtain in this framework, I present standard errors from a pairs bootstrap that was clustered on date of birth.

Looking to the first row of Panel A, it is clear that test age and schooling have strong effects on grade repetition, while entry age has weak effects. A one-year increase in age at the beginning of grade 3 reduces the probability of repeating by 8.6 percentage points, whereas a one-year increase in schooling reduces that probability by about 2 percentage points, and increases in entry age have a zero effect. Recall that the combined effect of entry age and test age estimated by the literature (α_1 in Table 2.6) is 7.8 percentage points, meaning that test age accounts for all of this effect, and entry age actually reduces the magnitude slightly.

The following two rows of Panel A separate the estimates by gender. It is still the case that a one-year increase in both test age and schooling have strong effects on grade repetition, but for males the magnitude of both effects is approximately double that of females. For males, a one year increase in test age leads to an 11.3 percentage point reduction in the probability of repeating a grade, a one-year increase in time spent in kindergarten leads to a 2.4 percentage point reduction, and interestingly a one-year increase in entry age leads to a 2 percentage point increase.

If one interprets differences due to test age as differences in maturity at the time a test is written or at the time a student begins a grade level, then because maturity in grade 3 plays such a large role in determining whether a student progresses to the next grade, the results make sense. If there is some minimum maturity standard that students must meet to satisfy the requirements to move ahead to the next grade, then the gender differences can be explained by males being slightly behind females on the maturational scale. Allowing males to be one year older would benefit them more since it would allow more of them to reach this minimum standard, whereas females would benefit less since many of them may have already reached that standard.

When test scores are the outcome, Panels B and C show that the story is entirely different. For both numeracy and reading scores, entry age has strong effects, while schooling and test age have very weak effects. A one-year increase in entry age increases both numeracy and reading scores by roughly 0.071σ, whereas increases in test age and schooling have much smaller, statistically insignificant effects. Entry age thus accounts for almost all of the total effect α_1. These results are at odds with Black et al. (2008) and with Crawford et al.

(2007), who find that most of the differences in test scores are due to test age, which could be related to my use of BC data versus their use of Norwegian and English data.

Separating the estimates by gender highlights even more interesting results. It appears as though the entry age effect is almost entirely driven by males, whose numeracy and reading scores increase by 0.122σ and 0.081σ with every one-year increase in entry age. The comparable effect for females is small, and always statistically insignificant. Both genders fail to show a significant association between test age and test scores. Finally, the effect of one year of extra schooling on female test scores is large and negative at about -0.121σ for reading and -0.078σ for numeracy, whereas for males this effect is close to zero and statistically insignificant.

The results suggest that cognitive ability is strongly related to the age at which students enter school, independent of age differences that exist at the time test are taken. Unlike grade repetition, it appears that accumulated skill is playing a large role in terms of student performance on tests. If there is some expected standard that students must reach to be ready to learn when they enter school, and males' brains development is slightly behind that of females, then by allowing an extra year before entry, more males would be able to reach this minimum threshold, and this would explain the larger entry age effect for males. The effect of one year of extra schooling on female test scores is somewhat surprising, but it could be related to kindergarten being a poor substitute for what females would be doing if they had not stayed in kindergarten for that long. For example, perhaps females benefit much more from spending time with their parents, or in market-based child care, and that taking this away from them has made them perform worse on tests all the way into grade 10.

Before discussing the policy implications of these conclusions, it is important to note that estimates in Table 2.8 are robust to the inclusion of baseline variables and school fixed effects. In this mode, if adding variables to the specification does not change the estimated coefficients, then this is further evidence that no birth date targeting has taken place. Indeed, as every second column in Table 2.8 shows, the coefficients are largely insensitive to the inclusion of baseline characteristics and school fixed effects. The only estimates that do show some change are the female schooling estimates, which decrease in size and lose statistical significance. The main entry age and test age results are, however unaffected by the inclusion of these variables, and actually become more precise. The available evidence thus points to a very robust set of estimates.

2.9 Conclusions and Policy Implications

Disentangling the entry age from the test age effect is important for parental behaviour and policy. Most of the existing research that estimates the relationship between age and student outcomes has not been able to separate these distinct effects. In this paper I solve this identification problem by comparing the outcomes of children enrolled in BC schools around the time of the dual entry experiment. By decoupling entry age from test age, this policy allowed me to provide estimates of the two independent effects on grade repetition and test scores, while also allowing me to estimate the effect of one year of extra kindergarten.

I reach several interesting conclusions. The first is that being one year older at the beginning of grade 3 greatly reduces the probability of repeating that grade, even among students who entered kindergarten at different times. Second, regardless of age differences that exist at the time children are tested, a student entering school one year later will face a long term advantage in both reading and numeracy ability. Third, both the test age effect on grade repetition and the entry age effect on test scores are much more pronounced for males. Finally, though spending extra time in kindergarten reduces the probability that a student of either gender repeats a grade, it is not necessarily beneficial for cognitive ability, and can even lead to worse performance in the case of females.

How can these results help inform parental decisions and policy? From a parental point of view, there appears to be a substantial benefit to delaying a child's entry into school, especially if that child is male. Being older at entry will ensure that the child is ready to learn, and lead to higher test scores. At the same time, delaying entry makes the student one year older at every stage in the education system, and would therefore reduce the probability that they repeat a grade, at least in early elementary school. If delayed entry is costly in terms of shifting time away from market work and towards caring for the child, or even in terms of the costs of centre-based child care, another option would be some sort of compensatory investment in younger children. Extra books, time, or after-school programs would likely benefit the younger students and help bring them towards the level of the older students at entry.

From a policy perspective, correcting test score imbalances between younger and older students requires that they be targeted when they enter school, since it can last at least until grade 10 and possibly beyond. An example of such a policy would be raising the maximum legislated entry age, which could have the effect of increasing cognitive ability all the way out to grade 10 and beyond. Schools could also provide compensatory investments in younger children, with extra instructional time or teaching assistants targeted towards the youngest students in the class at entry. It is important that such investments keep gender in mind, however, since the results from this paper show that younger males suffer worse than younger females from being one year younger in the classroom.

It is interesting to note that grade repetition as a mechanism to help students catch up to their peers appears to be completely ineffective for both genders according to estimates from this paper. In essence, grade repetition is a combination of one extra year of schooling and making the student one year older at the beginning of every grade. I show in the results that such a combination yields zero benefit in terms of both numeracy and reading test scores in grade 10. In addition, repeating a grade may have negative consequences on outcomes not measured in this paper, perhaps of an emotional nature. Given no long run academic benefit and what appear to be substantial costs, grade repetition in does not seem like a good strategy for improving the outcomes of students who are lagging behind.

An issue that has received a great deal of attention in the media recently is the poor performance of boys relative to girls along many school outcomes, which include international standardized tests such as the Programme for International Student Assessment (PISA), higher dropout rates, lower probability of attending post-secondary education. The results given in this paper imply that targeting corrective policy towards younger students may help reduce the gender gap in performance since boys benefit so much more than girls in terms of increases in entry age and test age. Indeed, Sax (2007) continues to advocate that boys delay entry one year past the legislated entry date so that they are as ready to learn as girls and would therefore be operating on the same level all the way through school.

References

Bedard, Kelly, and Elizabeth Dhuey (2006) 'The persistence of early childhood maturity: International evidence of long-run age effects.' *The Quarterly Journal of Economics* 121(4), 1437–1472

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes (2008) 'Too young to leave the nest: The effects of school starting age.' Working Paper 13969, National Bureau of Economic Research, March

Card, David, and David S. Lee (2006) 'Regression discontinuity inference with specification error.' Number 322. In 'NBER Technical Working Paper.'

Cascio, Elizabeth, and Diane Whitmore Schanzenbach (2007) 'First in the class? age and the education production function.' Number 13661. In 'NBER Working Paper.'

Crawford, Claire, Lorraine Dearden, and Costas Meghir (2007) 'When you are born matters: The impact of date of birth on child cognitive outcomes in england.' Technical Report, The Institute for Fiscal Studies, October

Datar, Ashlesha (2006) 'Does delaying kindergarten entrance give children a head start?' *Economics of Education Review* 25(1), 43–62

Deming, David, and Susan Dynarski (forthcoming) 'The lengthening of childhood.' *The Journal of Economic Perspectives*

Dhuey, Elizabeth, and Stephen Lipscomb (2007) 'Disabled or young? relative age and special education diagnoses in schools.' Mimeo: University of Toronto

Elder, Todd E., and Darren H. Lubotsky (2007) 'Kindergarten entrance age and children's achievement: Impacts of state policies, family background, and peers.' University of Illinois at Urbana-Champaign and Michigan State University

Fredriksson, Peter, and Björn Öckert (2006) 'Is early learning really more productive? the effect of school starting age on school and labor market performance.' Number 2006:12. In 'The Institute for Labour Market Policy Evaluation.'

Gammage, Philip (1991) 'Changing ideologies and provision in western canadian primary education.' *Comparative Education* 27(3), 311–323

Kawaguchi, Daiji (2006) 'The effect of age at school entry on education and income.' Number 162. In 'ESRI Discussion Paper.'

McEwan, Patrick, and Joseph Shapiro (2008) 'The benefits of delayed primary school enrollment: Discontinuity estimates using exact birth dates.' *The Journal Of Human Resources* 93(1), 1–29

Puhani, Patrick A., and Andrea M. Weber (2006) 'Does the early bird catch the worm? instrumental variable estimates of educational effects of age of school entry in germany.' Diskussionspapiere der Wirtschaftswissenschaftlichen Fakultät der Universität Hannover dp-332, Universität Hannover, Wirtschaftswissenschaftliche Fakultät, April

Sax, Leonard (2007) *Boys Adrift: The Five Factors Driving The Growing Epidemic of Unmotivated Boys and Underachieving Young Men* (Basic Books)

Stock, James H., and Motohiro Yogo (2002) 'Testing for weak instruments in linear iv regression.' NBER Technical Working Papers 0284, National Bureau of Economic Research, Inc, November

Table 2.1: Grade 3 and 10 Sample Exclusions

	Grade 3 Sample		Grade 10 Sample	
	Students	%of Total	Students	% of Total
Universe	166,266	100	145,765	100
Exclude:				
Missing Test Score	3,396	2	5,053	3
Not observed all years	1,279	1	2,880	2
Final Sample	161,591	97	137,832	95

Sample universe in each column includes the first observation of all students who are observed at entry, born between 1984 and 1987 inclusive. Note that the percentage column is sensitive to the ordering of the exclusions, but would be only slightly different if they were performed in a different order.

Table 2.2: Average Student Characteristics

Cutoff	All	Jan 1,'85		Nov 1,'85		Jan 1,'86		May 1,'86		Jan 1,'87	
Months From Cutoff		-1	+1	-1	+1	-1	+1	-1	+1	-1	+1
Panel A: Grade 3											
Entry Age	5.15	5.16	5.62	4.92	5.15	5.12	4.96	4.72	5.30	4.83	5.63
Test Age	8.19	8.18	8.61	7.90	7.92	7.91	8.55	8.36	8.30	7.83	8.63
Repeat Grade	2.30	2.12	0.75	3.50	5.43	6.32	0.74	1.44	1.40	7.08	0.93
Male	51.35	51.26	50.23	51.07	51.40	52.39	52.32	55.21	50.63	50.90	52.40
Aboriginal	9.52	9.43	9.64	9.77	10.02	9.58	10.32	9.08	9.37	10.51	11.53
Non-English	11.52	11.44	12.44	11.62	12.79	12.88	11.70	9.92	10.98	11.46	13.01
Panel B: Grade 10											
Entry Age	5.15	5.15	5.62	4.92	5.14	5.11	4.96	4.72	5.30	4.82	5.63
Age	15.89	15.89	16.29	15.63	15.65	15.66	16.23	16.03	15.98	15.58	16.33
Numeracy Score	0.07	0.07	0.08	0.01	0.03	0.05	0.06	0.11	0.14	0.06	0.10
Reading Score	0.15	0.15	0.15	0.10	0.11	0.10	0.14	0.18	0.21	0.14	0.20
Male	50.78	50.64	49.59	50.50	51.24	51.79	52.33	54.69	50.55	50.46	51.62
Aboriginal	7.90	7.78	7.79	8.02	8.72	7.87	9.05	7.73	8.30	9.07	9.44
Non-English	12.08	12.01	13.04	11.99	13.18	13.44	12.29	10.26	11.47	11.95	14.08

+1 indicates students are born one month after the listed cutoff, whereas -1 indicates they are born one month before the cutoff. Age is in years; Repeat Grade, Male, Aboriginal, Non-English are all percentages; Numeracy and Reading Scores are z-scores with mean 0 and standard deviation 1. All refers to the entire sample of students, and the dates listed in the header refer to a school entry cutoff point.

Table 2.3: School Start Times for Students Born 1984-1987

Birth Date	Expected Start Date	Total	%On Time	%Late	%Early
Panel A: All Students					
Jan - Dec 84	Sep-89	40265	96.56	2.75	0.69
Jan - Oct 85	Sep-90	35525	93.74	5.72	0.55
Nov - Dec 85	Jan-91	6623	49.00	21.83	29.17
Jan - Apr 86	Jan-91	11587	98.59	0.22	1.19
May - Dec 86	Sep-91	27100	97.37	2.54	0.09
Jan - Dec 87	Sep-92	40491	98.45	1.46	0.06
Panel B: Males					
Jan - Dec 84	Sep-89	20614	95.87	3.58	0.55
Jan - Oct 85	Sep-90	18248	93.40	6.24	0.36
Nov - Dec 85	Jan-91	3428	46.14	26.48	27.38
Jan - Apr 86	Jan-91	6178	98.79	0.26	0.96
May - Dec 86	Sep-91	13748	96.32	3.59	0.09
Jan - Dec 87	Sep-92	20757	98.03	1.93	0.03
Panel C: Females					
Jan - Dec 84	Sep-89	19651	97.28	1.89	0.83
Jan - Oct 85	Sep-90	17277	94.10	5.16	0.74
Nov - Dec 85	Jan-91	3186	52.07	16.82	31.10
Jan - Apr 86	Jan-91	5409	98.37	0.17	1.46
May - Dec 86	Sep-91	13352	98.46	1.46	0.08
Jan - Dec 87	Sep-92	19734	98.96	0.96	0.08

Expected start date is derived from BC law, and is based on when the student turns five years old. On time refers to the percentage of students who begin at the assigned entry date, late refers to students who postpone entry by one or two entry dates, and early refers to students who begin one or two entry dates before their normal entry date.

Table 2.4: Average Baseline Characteristics of Compliers at Cutoffs

Cutoff	Male	Aboriginal	Non-English
Jan 1, '85	50.1	9.0	13.3
Nov 1, '85	49.5	9.3	12.1
Jan 1, '86	51.5	9.9	13.1
May 1, '86	52.7	9.2	10.4
Jan 1, '87	50.2	11.1	12.7

The numbers generated in this table use only the sample of compliers born within 1 month of each of the five cutoff points listed in the table. Non-English refers to the student's home language being anything other than English. Cutoff refers to the legislated school entry deadline points.

Table 2.5: First Stage Estimates of Age Differences at Cutoffs

Entry Cutoff	Grade Repetition		Test Score	
	Test Age	Entry Age	Test Age	Entry Age
cut_{J85i}	0.795***	0.822***	0.716***	0.830***
	(0.008)	(0.008)	(0.011)	(0.008)
cut_{N85i}		0.296***		0.288***
		(0.009)		(0.009)
cut_{J86i}	0.650***		0.574***	
	(0.009)		(0.011)	
cut_{M86i}		0.639***		0.642***
		(0.006)		(0.006)
cut_{J87i}	0.856***	0.837***	0.811***	0.852***
	(0.007)	(0.008)	(0.009)	(0.008)
R^2	0.692	0.713	0.492	0.776
F-Stat	8935	7202	4349	7618
N	161591	161591	137832	137832

*p<0.10, **p<0.05, ***p<0.01. $cut_{J85i} = 1\{bd_i >=$ Jan 1, 1985$\}$, $cut_{N85i} = 1\{bd_i >=$ Nov 1, 1985$\}$, $cut_{J86i} = 1\{bd_i >=$ Jan 1, 1986$\}$, $cut_{M86i} = 1\{bd_i >=$ May 1, 1986$\}$, and $cut_{J87i} = 1\{bd_i >=$ Jan 1, 1987$\}$. Estimates are in years. Grade repetition refers to the grade 3 sample, while test score refers to the grade 10 sample. Standard errors are clustered on birth date. F-Stat is a joint test of significance for all the variables listed in the Table.

Table 2.6: Regression Discontinuity Estimates of Entry Age and Test Age Effects

	Entry Age+ Test Age (α_1)	Test Age+ Schooling (α_2)	Entry Age+ Schooling (α_3)
Panel A: Gr 3 Repetition			
Estimate	-0.078***	-0.105***	-0.011
	(0.004)	(0.009)	(0.010)
R^2	0.012	0.009	0.011
Under-id F-Stat	23129	13162	20779
Weak-id F-Stat	17970	12617	14674
N	161591	161591	161591
Panel B: Gr 10 Numeracy			
Estimate	0.082***	-0.035	0.028
	(0.026)	(0.055)	(0.054)
R^2	-0.006	0.004	0.001
Under-id F-Stat	19481	11971	18241
Weak-id F-Stat	11396	9179	13407
N	137832	137832	137832
Panel C: Gr 10 Reading			
Estimate	0.091***	0.009	0.060
	(0.026)	(0.052)	(0.050)
R^2	-0.007	0.001	0.002
Under-id F-Stat	19481	11971	18241
Weak-id F-Stat	11396	9179	13407
N	137832	137832	137832

*p<0.10, **p<0.05, ***p<0.01. α_1 is the joint entry age and test age effect, α_2 is the joint test age and schooling effect, and α_3 is the join entry age and schooling effect. The critical values for the underidentification and weak identification statistics can be found in Stock and Yogo (2002). Standard errors are clustered on birth date.

Table 2.7: Discontinuities in Baseline Characteristics

	Male	Aboriginal	Non-English	Male	Aboriginal	Non-English
cut_{J85i}	-0.022	0.011	-0.005	-0.024	0.004	-0.012
	(0.015)	(0.009)	(0.010)	(0.016)	(0.008)	(0.011)
cut_{N85i}	0.002	0.003	-0.001	-0.002	0.002	-0.021
	(0.034)	(0.020)	(0.014)	(0.036)	(0.019)	(0.014)
cut_{J86i}	0.026	-0.001	-0.024	0.038	0.003	-0.041*
	(0.035)	(0.019)	(0.017)	(0.033)	(0.017)	(0.021)
cut_{M86i}	-0.056***	0.002	0.008	-0.058**	0.005	-0.009
	(0.021)	(0.013)	(0.012)	(0.024)	(0.013)	(0.010)
cut_{J87i}	0.003	0.009	0.029**	0.001	0.002	0.033***
	(0.016)	(0.011)	(0.012)	(0.017)	(0.011)	(0.011)
R^2	0.000	0.000	0.001	0.000	0.000	0.002
F-Stat	2.203	0.473	1.728	2.129	0.086	3.222
N	161591	161591	161591	137832	137832	137832

*p<0.10, **p<0.05, ***p<0.01. $cut_{J85i} = 1\{bd_i >= \text{Jan 1, 1985}\}$, $cut_{N85i} = 1\{bd_i >= \text{Nov 1, 1985}\}$, $cut_{J86i} = 1\{bd_i >= \text{Jan 1, 1986}\}$, $cut_{M86i} = 1\{bd_i >= \text{May 1, 1986}\}$, and $cut_{J87i} = 1\{bd_i >= \text{Jan 1, 1987}\}$. Each column is a different regression, where the dependent variable is a binary indicator defined by the column title. The F-stat is from a test of the null that all of the cut variables are jointly equal to zero against the alternative that they are not. Standard errors are clustered on birth date.

Table 2.8: Structural Entry Age, Test Age, and Schooling Estimates

	Entry Age (β_1)		Test Age (β_2)		Schooling (β_3)	
Panel A: Gr 3 Repetition						
All	.007	.008	-.086***	-.086***	-.019***	-.021***
	(.006)	.(006)	(.007)	(.006)	(.007)	(.004)
Male	.019**	.022**	-.113***	-.114***	-.024**	-.026**
	(.010)	(.010)	(.010)	(.010)	(.011)	(.011)
Female	-.004	-.006	-.058***	-.057***	-.013	-.018**
	(.007)	(.007)	(.008)	(.007)	(.009)	(.008)
Panel B: Gr 10 Numeracy						
All	.072*	.075**	.010	-.001	-.044	-.001
	(.039)	(.036)	(.042)	(.039)	(.041)	(.039)
Male	.122***	.136***	-.050	-.064	.030	.048
	(.042)	(.041)	(.057)	(.052)	(.042)	(.040)
Female	.020	.017	.071	.060	-.121***	-.058*
	(.045)	(.043)	(.051)	(.048)	(.045)	(.044)
Panel C: Gr 10 Reading						
All	.071*	.067**	.020	.019	-.011	.020
	(.036)	(.033)	(.039)	(.037)	(.039)	(.037)
Male	.081**	.100**	.019	.005	.035	.054
	(.045)	(.044)	(.090)	(.079)	(.050)	(.048)
Female	.038	.038	.040	.038	-.078*	-.022
	(.045)	(.043)	(.072)	(.065)	(.046)	(.044)
Controls						
Baseline	No	Yes	No	Yes	No	Yes
School FE	No	Yes	No	Yes	No	Yes

*p<0.10, **p<0.05, ***p<0.01. Baseline variables include an indicator for male, ever Aboriginal and ever home language other than English. Standard errors are bootstrapped 999 times using a pairs bootstrap.

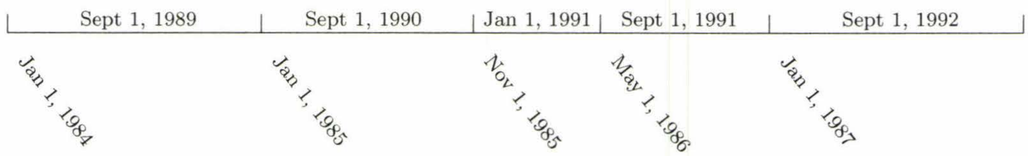


Figure 2.1: Birth Dates (bottom) and Corresponding Entry Dates (top) in BC 1989 - 1992

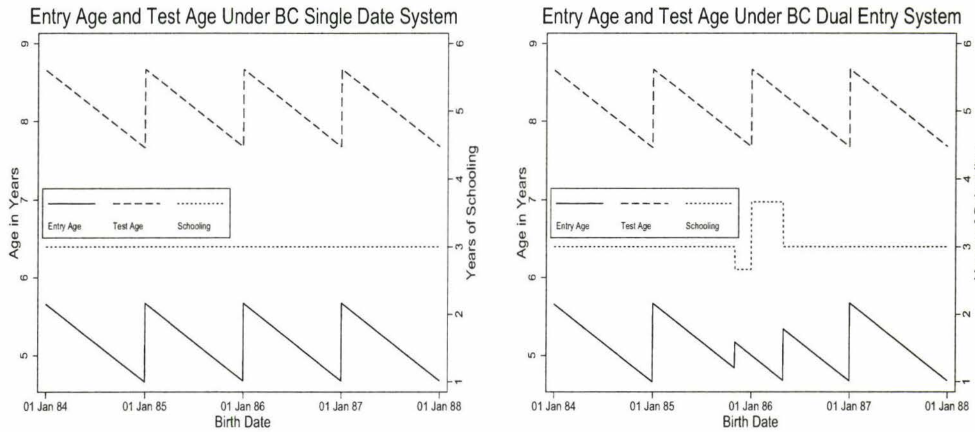


Figure 2.2: Entry Ages Under Single and Dual Entry in BC

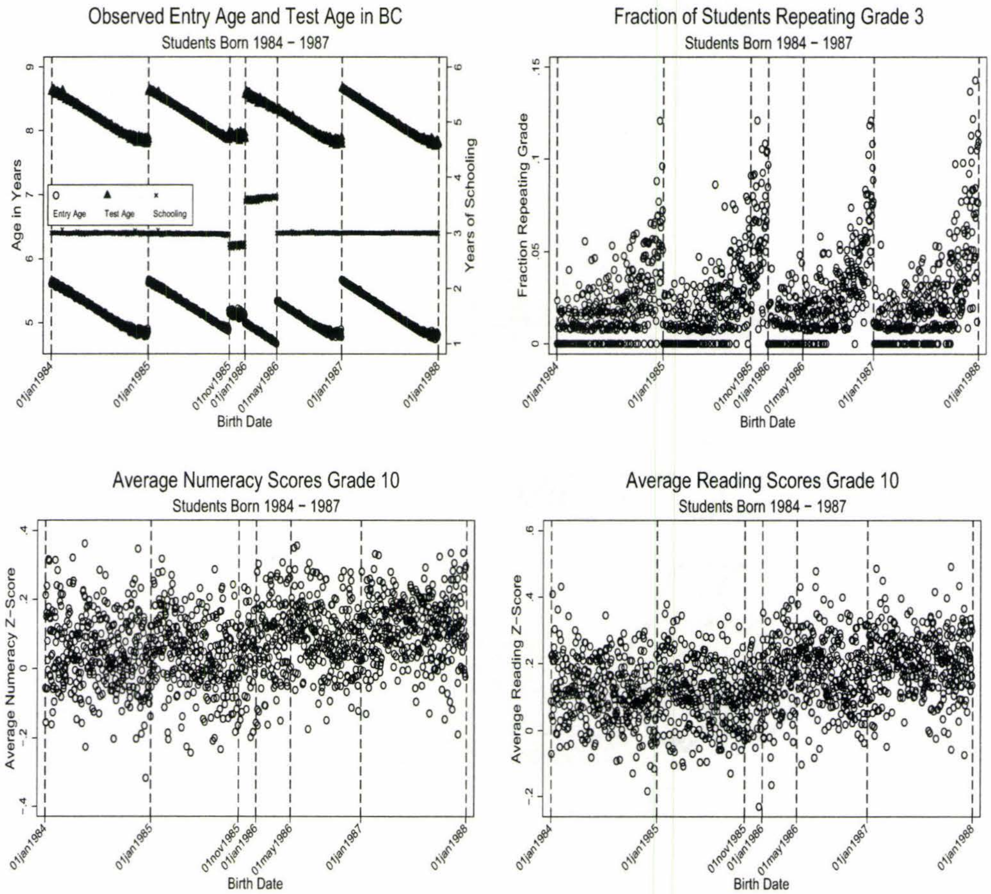


Figure 2.3: Average Entry Age, Test Age, Repetiton, and Test Scores by Birth date

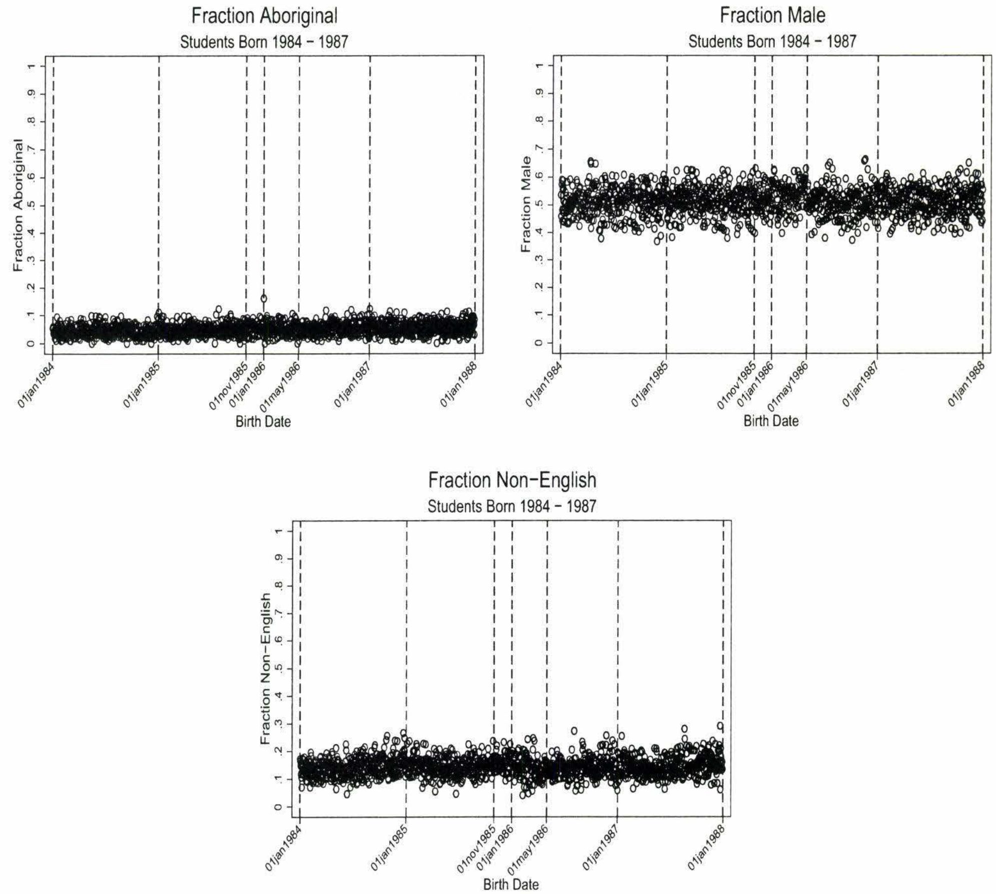


Figure 2.4: Average Baseline Measures by Birth date

2.A Appendix: Tables by Gender

Table 2.A1: First Stage Estimates of Age Differences at Cutoffs by Gender

	Grade Repetition		Test Score	
	Test Age	Entry Age	Test Age	Entry Age
Panel A: Males				
cut_{J85i}	0.756***	0.784***	0.686***	0.794***
	(0.011)	(0.011)	(0.015)	(0.011)
cut_{N85i}		0.315***		0.308***
		(0.014)		(0.015)
cut_{J86i}	0.624***		0.526***	
	(0.011)		(0.016)	
cut_{M86i}		0.630***		0.630***
		(0.008)		(0.008)
cut_{J87i}	0.798***	0.771***	0.729***	0.784***
	(0.009)	(0.011)	(0.014)	(0.012)
R^2	0.653	0.672	0.446	0.736
F-Stat	4371	3548	1739	3277
N	82982	82982	69985	69985
Panel B: Females				
cut_{J85i}	0.839***	0.864***	0.752***	0.869***
	(0.012)	(0.013)	(0.015)	(0.012)
cut_{N85i}		0.276***		0.267***
		(0.011)		(0.012)
cut_{J86i}	0.676***		0.625***	
	(0.011)		(0.015)	
cut_{M86i}		0.650***		0.655***
		(0.006)		(0.005)
cut_{J87i}	0.917***	0.905***	0.896***	0.923***
	(0.007)	(0.009)	(0.011)	(0.006)
R^2	0.740	0.764	0.551	0.824
F-Stat	7166	6159	3368	10578
N	78609	78609	67847	67847

*p<0.10, **p<0.05, ***p<0.01. $cut_{J85i} = 1\{bd_i >= \text{Jan 1, 1985}\}$, $cut_{N85i} = 1\{bd_i >= \text{Nov 1, 1985}\}$, $cut_{J86i} = 1\{bd_i >= \text{Jan 1, 1986}\}$, $cut_{M86i} = 1\{bd_i >= \text{May 1, 1986}\}$, and $cut_{J87i} = 1\{bd_i >= \text{Jan 1, 1987}\}$. Estimates are in years. Grade repetition refers to the grade 3 sample, while test score refers to the grade 10 sample. Standard errors are clustered on birth date.

Table 2.A2: Regression Discontinuity Estimates of Entry Age and Test Age Effects by Gender

	Males			Females		
	Entry Age+ Test Age (α_1)	Test Age+ Schooling (α_2)	Entry Age+ Schooling (α_3)	Entry Age+ Test Age (α_1)	Test Age+ Schooling (α_2)	Entry Age+ Schooling (α_3)
Panel A: Gr 3 Repetition						
Estimate	-0.095***	-0.137***	-0.005	-0.062***	-0.072***	-0.017
	(0.007)	(0.013)	(0.014)	(0.005)	(0.011)	(0.011)
R^2	0.017	0.012	0.015	0.007	0.006	0.007
Under-id F-Stat	11394	6882	10553	11860	6345	10302
Weak-id F-Stat	8022	6556	7880	10529	6110	6801
N	82982	82982	82982	78609	78609	78609
Panel B: Gr 10 Numeracy						
Estimate	0.072	-0.019	0.152**	0.100**	0.055	-0.101
	(0.045)	(0.092)	(0.076)	(0.041)	(0.079)	(0.081)
R^2	-0.005	0.003	0.000	-0.009	-0.004	0.003
Under-id F-Stat	9603	6095	9193	9603	6095	9133
Weak-id F-Stat	5436	4657	7078	5436	4657	6340
N	69985	69985	69985	69985	69985	67847
Panel C: Gr 10 Reading						
Estimate	0.091***	-0.050	0.118*	0.078**	-0.038	-0.040
	(0.032)	(0.066)	(0.071)	(0.035)	(0.066)	(0.075)
R^2	-0.006	0.006	0.001	-0.004	0.007	0.004
Under-id F-Stat	9977	5941	9193	9977	5941	9133
Weak-id F-Stat	6073	4565	7078	6073	4565	6340
N	67847	67847	69985	67847	67847	67847

*p<0.10, **p<0.05, ***p<0.01. α_1 is the joint entry age and test age effect, α_2 is the joint test age and schooling effect, and α_3 is the join entry age and schooling effect. The critical values for the underidentification and weak identification statistics can be found in Stock and Yogo (2002). Standard errors are clustered on birth date.

Chapter 3

Learning About School Quality: Does New Information Affect School Choice?

3.1 Introduction

Proponents of public school choice argue that by inducing competition between schools for students, school choice can create the incentive for schools to produce education more efficiently. Underlying this assertion is the notion that parents will respond to poor school performance by choosing a better one, forcing inefficient schools to improve or lose education dollars. A key assumption is that most parents are well-enough informed to know which schools are good and which schools are bad. The fact that recent school accountability frameworks such as the No Child Left Behind Act (NCLB) and the Florida A+ Education Plan focus in part on grading schools and reporting their progress to the public suggests that parents are not well informed about school quality. Without adequate information, even rational parents may not be selecting schools that produce the best outcomes for their children, reducing the competitive pressures faced by schools to keep quality high. Policies that promote the dissemination of better information should lead to a better matching of students with schools, and stronger student performance.

This paper aims to explain how new information affects school choice, and whether there is empirical support for this explanation. I use a Normal Learning Model (NLM) embedded inside a utility maximization framework to obtain conditions under which parents choose a different school after gaining insights on school quality. The conclusions are that parents would use the new information to choose a different school if they had poor prior information on school quality, if the new information is not too noisy, and if the new information does not simply confirm what was already known. I show further that when parents have several observations on the new information, they shift weight away from the old information and towards the new because the multiplicity of observations allows them to better identify true school quality. Finally, I show that when some groups of parents have less-precise prior school quality information, they put more weight on the new information relative to parents with more-precise priors. For them, the test score represents a better indicator of true school quality.

I test this framework using a natural experiment from the province of British Columbia (BC), Canada. In 1999 the BC government started releasing school level standardized test score information. In the main analysis, I study students from the Lower Mainland of BC and evaluate whether the probability of switching out of a school varies with test scores after the information was revealed. The results suggest that all else equal parents are moderately less likely to leave a school that scores well on the test, however as time passes and more test scores are revealed, the negative relationship between test scores and leaving a school is stronger, supporting one of the main predictions of the model that parents put more weight on new information over time. Finally, I show this response varies across the education level of the student's neighbourhood and across language spoken at home.

A major concern with the main analysis is that the muted response to test scores might be caused by constraints on mobility. For the first three years that test scores were observed by parents, students were required to attend the school in their catchment area.[1] Thus, to switch schools, parents had to move to a different neighbourhood, which carries a substantial cost. I address this issue in two ways. First, in 2003 BC implemented legislation called "open boundaries" that allowed children to attend any school in the province provided there was enough space.[2] I examine the extent to which this lowered moving costs by checking whether the response to test scores is stronger in the years following open boundaries. The results show that the negative relationship between test scores and the probability of leaving a school gets stronger.

I also determine the role of moving costs by studying kindergarten enrollment. The rationale behind this strategy is that parents should value moving

[1] Catchment area is the set of households surrounding a school that defines which students are allowed to enroll in that school.

[2] The legislation contains a very precise priority ordering for students to be able to attend a particular school, with students living in the neighbourhood being first on the list. Thus, living in the neighbourhood is still the best way to guarantee your attendance at a preferred school.

costs less in their initial school choice, since making a good choice now will save them from having to switch schools in the future. Thus, observing unusually large flows of kindergarten enrollment into schools scoring well on the test would indicate moving costs are playing a role in the small size of the main effect. The results of this analysis show that enrollment does indeed flow into schools scoring well on the test, but the size of the effect is not so large that one would expect the mobility in the upper grades to be constrained by moving costs.

I supplement evidence on kindergarten enrollment with a discrete choice analysis of individual-level school choices. Unlike the school-level enrollment regressions, the discrete choice analysis can shed light on heterogeneity in student responses to test scores across student characteristics. I apply the Conditional Logit model to school choices in kindergarten and show that the probability of choosing a school depends positively on test scores. For the average individual, a 1 standard deviation (σ) increase in test scores has the same effect on preferences as an 8.76 percentage point decrease in the proportion Aboriginal, 56.52 percentage point decline in the unemployment rate, a \$12,380 increase in the income of the neighbourhood around the school, and a 0.16 km decrease in distance from school. Furthermore, the strength of the test score effect depends negatively on whether the individual speaks English at home and on the proportion of people living in their neighbourhood that have low education.

The results suggest a small effect of the information release on student mobility between schools that depends in particular on the proportion of the neighbourhood that is poorly educated and whether an individual speaks English at home. The small size may be because parents care less about the sorts of school attributes that test scores capture (such as class size or peer quality) than they do about attributes captured by neighbourhood demographics, or because adding a small piece to the myriad of information that already exists on school quality is simply not enough to have a large impact of their decisions. Since a large amount of resources are devoted to administering and publishing the results of such tests, the results imply that at least when it comes to their actual use for school selection, the costs may outweigh the benefits.

3.2 Previous Literature

This study relates firstly to a strand of literature that estimates parental preferences for school quality within a school or residential choice framework. One approach to this problem is to look at the observed choices of parents and check whether they are correlated with measures of school quality, most frequently measured as test scores. Hastings et al. (2005) uses data taken from school choice forms in the Charlotte-Mecklenburg school district (CMS) to evaluate what weight parents place on test scores versus other commonly observed characteristics of the school. Unlike many other jurisdictions, CMS has a full school choice program in place, so it is conceivable that parents will have a large degree of latitude when exercising choice. They find that parents do care about test scores, and that this preference is increasing in the income of parents. The weight place on school test scores relative to other measures such as distance and peer characteristics, however, appears to be low. Barrow (2002) compares the residential choices made by families with and without children under 18 to show that white households would be willing to pay about \$1800 more for a house for each 100 point increase in SAT test scores. Finally, Bayer et al. (2004) provides a marginal willingness to pay for test scores at about 2% per 1 standard deviation increase in test scores, and notes that this direct effect is likely small because it does not take into account the resorting by peer characteristics that occurs when households relocate due to changes in test scores.

A second approach to estimating preferences for school quality that is commonly used when no direct information is available on parent choices is to examine the relationship between house prices and test scores. The reasoning behind this approach is that if parents care about school quality, it should be capitalized into the price of the house. Black (1999) uses a boundary fixed effects procedure to control for the relationship between neighbourhood characteristics and test scores, and shows that households are willing to pay 2.5% more for a house located in an area with 5% better test scores. Similarly, Gibbons and Machin (2003) show that a 1% increase in the proportion of students meeting education target grades increases property prices by 0.67%.

Another strand of related literature attempts to estimate what happens when the cost of obtaining information on school quality is reduced. Hastings and Weinstein (2007) examine the effect of information sheets distributed to students attending delinquent schools as defined by the American No Child

Left Behind (NCLB) legislation.[3] According to this law, students who attend schools that receive federal funding under the Title I program, and that fail to make adequate yearly progress, are entitled to switch to a different non-NCLB school. In addition to their entitlement to switch, they receive an information sheet indicating the quality of all schools in the district to help with their choice. Hastings and Weinstein (2007) examine the effect of this choice coupled with the information shock on parental preferences, and find that 16% of those who were entitled to switch actually exercise their option. Furthermore, those 16% of students chose schools with test scores 1σ higher than the school they chose a few months earlier. Effects were heterogeneous, with black, high ability students, attending magnet schools, living far from their NCLB school, and living in areas with high scoring schools all being more likely to exercise choice.

In a related study, Hastings et al. (2007) distribute simplified information sheets randomly to students in NCLB and to low- to middle-income neighbourhood schools to examine the effect on parental preferences for school quality. In their experiment, they distribute two kinds of sheets: one containing average test scores of all schools, and one average test scores *and* odds of entry into those schools. The effect of information sheets including only information about test scores doubles the estimated preference parameter on school test scores, an effect which is not heterogeneous by income level. The information sheets containing both average test scores and odds of admission amplify the effect for high income students, while dampening it for low income students.

[3]for more information on NCLB in this context, see the discussion in Hastings and Weinstein (2007).

Hussain (2007) uses similar ratings data from England to estimate the effect of information on enrollment levels at schools of varying qualities. The crux of his identification strategy is that not all schools are rated at the same time, so at moments when one school is being rated, similar schools that are not rated provide natural controls. He shows that schools rated "fail" see reduced enrollment of up to 6% three years after the ratings were released, while schools rated "very good" see an increase of up to 2%. Interestingly, the responses were similar across schools located in more and less deprived areas.

Using the house price approach, Figlio and Lucas (2004) studies whether simplified school grades provided though the Florida A+ education plan are capitalized into housing prices during the implementation and subsequent years of the accountability plan. The grades consisted of a report card-like system where schools were given an A through F depending on their quality. They find that A schools are valued 19.5% higher than B schools in the implementation stage of the program, and that B schools are valued 15.6% higher than C schools. This effect diminishes quickly in the following years, an effect they attribute to the noise inherent in this grading system.

Finally, there are a series of papers in health economics that look at the effect of information shocks on health insurance plan choice. The research design behind these studies takes advantage of years before and after information sheets were distributed to employees choosing their health insurance plan through the employer. The results imply that individuals tend to choose plans with better ratings.[4]

[4]See Scanlon et al. (2002), Beaulieu (2002), Chernew et al. (2001), Jin and Sorensen (2006), Pope (2006), Wedig and Tai-Seale (2002).

My contribution is to consider what happens when parents are given an entirely new source of information on school quality. In particular, while standardized test scores are routinely compiled and made publicly available in the United States, this is only a new phenomenon in Canada. Thus, parents are not accustomed to using test scores for school choice, and for them it is truly a new source of information that should have an impact on their decisions if they place value on them. Another novelty is that I provide a simple model of the theoretical effect that increased information might have, and use the data to test the predictions of that model.

3.3 The Effect of New Information on School Choice

3.3.1 Basic Model

Choosing a school is difficult because there are high costs to determine true quality. For example, a well-informed parent might need to make a visit to see the principal, ask neighbours, and visit several facilities, among other things. Even then, an idiosyncrasy in the fit between child and school makes true quality difficult to observe and to define. In the absence of perfect information, parents must form an expectation about the quality of schools where they might enroll their child, basing it on all information gathered in their search, then decide on the best school given their expectation. In such a world, new pieces of information like test scores are potentially valuable because they reduce the cost of uncovering true quality.

Below is a highly simplified model of parent school choice taking account of the uncertainty in school quality discussed above.[5] The choice process is simplified to focus attention on the effects of the extra information, which is approximated by the Normal Learning Model (NLM). Assume parent i's satisfaction (U_{is}) is determined by school s's quality (q_s),

$$U_{is} = q_s + \epsilon_{is} \tag{3.1}$$

where $\epsilon_{is} \sim N(0, h_{is}^{-1})$. In this specification, the zero mean of ϵ_{is} means that the average individual's utility is entirely determined by school quality, whereas others' decisions may deviate randomly from the average individual, with h_{is}^{-1} describing the spread of those deviations. School quality q_s is interpreted as an index of the things parents normally care about when choosing a school, such as teacher experience, peer ability, the state of technology at the school, the quality of a sports program, and class size. It is a summary measure of a school's desirability, such that increasing quality increases parent satisfaction.

Parents cannot observe school quality perfectly, so they must rely on expected quality when making their choice. The NLM provides a simple and convenient way to model the guess and its effects on choice behaviour.[6] The fundamental idea of NLM in this context is that parents are assumed to have prior beliefs that the distribution of quality is $q_s \sim N(X_s'\beta, h_{qi}^{-1})$ for each

[5]The notation is adapted from Moretti (2008), where a similar model is used for an analysis of peer effects in movie consumption.

[6]For recent examples of the normal learning model, see Moretti (2008) Ichino and Moretti (2006), Erdem and Keane (1996), Altonji and Pierret (2001), Lange (2007), and Chernew et al. (2001). Some studies refer to it as the Bayesian Learning Model.

school, and use the expectation of this distribution, $X_s^{'}\beta$, as their guess. $X_s^{'}\beta$ is a combination of school characteristics that parents already observe, such as neighbourhood income or the demographic composition of the school. Parents are assumed to know the precision of school quality information (h_{qi}).[7] While this precision is constant over schools, it varies over personal characteristics of parents, which implies, for example, that new immigrants could have a less-precise idea of school quality than a native born individual who has lived in the area for many years.

The average parent will pick school s if

$$
\begin{aligned}
E[U_{is}|X_s^{'}\beta] &> E[U_{ik}|X_s^{'}\beta], \forall k \neq s \\
X_s^{'}\beta &> X_k^{'}\beta, \forall k \neq s
\end{aligned} \tag{3.2}
$$

Absent any additional information, the result is that parents are observed choosing the school with the highest level of the combination of the observable attributes.

Suppose in an initial period that parents observe a noisy signal (S_s) about every school's quality,

$$S_s = q_s + \eta_s \tag{3.3}$$

with $\eta_s \sim N(0, h_{\eta i}^{-1})$. The zero mean of the noise component (η_s) implies that test scores are an unbiased source of school quality information. The precision

[7]The precision parameter is the inverse of the variance. I use this notation for convenience.

of test scores as a signal ($h_{\eta i}$) is assumed to be known and to vary across individuals. For concreteness, assume that the source of this signal is based on standardized test scores aggregated to the school level. The variation in the precision parameter could result from the test being less-meaningful to some groups of people, such as immigrants who have recently entered the country and are unfamiliar with the school evaluation system.

After observing the signal, parents update their expectations about each school's quality using Bayes' rule and end up with the following posterior beliefs about average school quality,

$$E[q_s|X_s^{'}\beta, S_s] = \left(\frac{h_{qi}}{h_{\eta i} + h_{qi}}\right) X_s^{'}\beta + \left(\frac{h_{\eta i}}{h_{\eta i} + h_{qi}}\right) S_s \tag{3.4}$$

School quality is now a weighted average between the new and the old information, where the weights placed on each piece depend only on their relative precision. If test scores have high precision relative to the old information, then a greater weight is placed on test scores. If it is mostly noise, then only minimal weight is shifted to test scores, and parents continue to use their prior information to judge school quality.

Defining $\theta_i = \left(\frac{h_{\eta i}}{h_{\eta i}+h_{qi}}\right)$ as the precision weight placed on test scores, and rearranging terms in Equation 3.4 slightly, utility for the average individual now becomes

$$E[U_{is}|X_s^{'}\beta, S_s] = \theta_i(S_s - X_s^{'}\beta) + X_s^{'}\beta \tag{3.5}$$

Define $S_s^* = (S_s - X_s^{'}\beta)$ to be the test score "shock," which represents the information contained in the test scores after netting out the information that parents already know. They will now choose school s over school k if ,

$$\theta_i(S_s^* - S_k^*) \quad > \quad (X_k^{'} - X_s^{'})\beta, \forall k \neq s \tag{3.6}$$

Equation 3.6 produces the result that parents choose school s if the shock relative to school k is high enough that it outweighs the differences in other characteristics of k versus s. Conversely, parents will choose school s if its characteristics are good enough relative to k that it outweighs the shock differential. The information contained in this shock is precision-weighted, and is only useful if the test scores contain useful information ($\theta_i > 0$). Furthermore, since θ_i varies across the population, certain groups of people will attach a different weight to the shock.

3.3.2 Model Dynamics

This section extends the model to accommodate a series of signals rather than the one-time event described in the previous section. Suppose a series of unbiased school quality signals are observed, which are represented in exactly the same way as in Equation 3.3. For ease of interpretation, suppose that we are observing one of these signals in each of a series of time periods. For example, in each school year, as the results of a yearly standardized test are revealed to the public, generating one signal each year. Let T represent the total number of signals observed. Expected school quality in year T becomes

$$(3.7)\quad E[q_{is}|S_{s1}, S_{s2}, ..., S_{sT}, X_s^{'}\beta] = \left(\frac{h_{qi}}{Th_{\eta i} + h_{qi}}\right) X_s^{'}\beta + \left(\frac{Th_{\eta i}}{Th_{\eta i} + h_{qi}}\right) \bar{S}_s$$

Expected school quality after T signals is nearly identical to the case where there is one signal, except that the weight placed on the signal increases as T increases, and now it depends on an average of the signals observed up to T, $\bar{S}_s = \frac{1}{T}\sum_{t=1}^{T} S_{st}$. Defining the term $\theta_{iT} = \left(\frac{Th_{\eta i}}{Th_{\eta i}+h_{qi}}\right)$ to be the precision weight after T signals, the term $S_s^{**} = (\bar{S}_s - X_s^{'}\beta)$ to be the cumulative shock, and rearranging equation 3.7, parents choose school s over school k if

$$(3.8)\quad \theta_{i\tau}(S_s^{**} - S_k^{**}) \;>\; (X_k^{'} - X_s^{'})\beta, \forall k \neq s$$

The intuition behind this maximization condition is the same as before, except that parents react to the average over the full set of shocks rather than a single value. The shock is still only useful if it is precise enough to draw out quality information, $(\theta_{i\tau} > 0)$.

This model contains several interesting - and testable - implications. Given a single information shock, parents may alter their original school choice decision if the signal provides enough new information that makes an alternative school look better. This is conditional on two main factors: one, the shock must provide new information on school quality rather than simply confirming what parents already know, and two, parents must place positive weight on the shock, meaning that the signal cannot be too noisy, relative to the prior information.

A second testable implication results from the assumption that the prior and signal precisions vary across personal characteristics. If this is true, different groups of parents will weight the shocks differently, since the weights are functions of the precisions, and we should observe heterogeneity in the response to new information. A priori, the model does not say what sorts of people might have noisier priors than others, and this is left entirely an empirical question.

A final implication is that as more signals are observed parents are more likely to react to them because the signals become more reliable indicators of true quality.

3.4 Empirical Model

To test the predictions of the model, I rely on a natural experiment from the province of British Columbia (BC), Canada in the late 1990's. Prior to school year 1999-2000, the BC Ministry of Education conducted a set of standardized tests on called the Provincial Learning Assessment Program (PLAP), which tested students in different subject areas every year or couple of years. Though this testing program existed since the 1970's, the results were never made publicly available at the school level, making it impossible for parents to use the results when choosing a school. This was all changed in the 1999-2000 school year when the province revamped its standardized testing system to focus on tracking student and school results over time on a yearly basis, and to make information available to the public. The new testing regime, called the Foundation Skills Assessment (FSA), evaluates grade 4, 7, and 10 students in the areas of numeracy, reading, and writing at the end of every school year. Results from the test are compiled into detailed, descriptive documents that are available online at the Ministry of Education website.[8]

The key outcome of this change in standardized testing regimes is that information previously unavailable to parents was suddenly made available to them. Treating the FSA release as an information shock, I will test the predictions of the model by comparing school choices before and after the information was available. I have a set of data containing scores on the FSA test since its inception, linked to administrative records that track students across schools from 1991 - 2004. The primary empirical analysis uses this data to examine changes in the probability that a student switches voluntarily out of their current school as a function of the school's score on the FSA test. I begin by looking at the impact of the information release in the first year it was available, then at how the effect changes over time after the results of several years of the FSA test are observed. Finally, I explain heterogeneity in switching behaviour across a small set of personal characteristics. The empirical analysis is structured in this way to highlight the fact that these are the empirical counterparts that test the theoretical predictions. Given the relative importance placed on early education, and the lower cost of switching, the analysis focuses on the switching behaviour of kindergarten to grade 4 students.

The main empirical issue is that the student population is not randomly

[8]See http://www.achievebc.ca/choices/search_about.aspx for an online tool that allows anyone to access a report on any school in BC.

distributed across schools because parents of similar socio-economic characteristics sort themselves into neighbourhoods through their residential location choice. If these characteristics bear upon test scores, then without adequate controls for school and neighbourhood variables, the independent effect of the test scores will not be identified. For example, it is generally accepted that parental income has a positive effect on student test scores, so it would be no surprise to find that a school located in a high income neighbourhood outperforms one in a low income neighbourhood, and without controlling for income, the effect of the test score would be confounded. Visually, this can be seen in Figure 1, where the map shows the location of schools colour-coded by their test score tercile on top of the income terciles of neighbourhoods. It is clear that schools doing well on test scores are located in good neighbourhoods.

To address this problem, I use a difference-in-differences-like approach that compares the probability of switching as a function of test scores before and after the information was available. The main estimating equation takes the form

$$\begin{aligned} (3.9) \qquad switch_{isgt} &= \alpha_0 + \alpha_1 score_{st} + \alpha_2 d^{after} + \alpha_3 score_{st} \times d^{after} \\ &+ Z_{it}^{'}\gamma + X_{sgt}^{'}\beta + \delta_s + \epsilon_{isgt} \end{aligned}$$

where variaton is over (i)ndividuals, (s)chools, (g)rades, and (t)ime. The variable $switch_{isgt}$ is an indicator equal to 1 if the student switches schools voluntarily between t and $t+1$; $score_{st}$ is the average combined score of the reading and numeracy FSA tests in grade 4 of year t; d^{after} is a dummy equal to

1 if the student is observed on or after the date of the release of the information; Z_{it} is a vector of time-constant and time varying student characteristics; X_{sgt} are time-varying school characteristics; and δ_s are a set of school fixed effects.[9] The variables contained within those vectors are discussed in more detail in the data section.

In this specification the coefficient of interest is α_3, which measures the change in the relationship between test scores and switching after the new information is revealed. The expected sign of α_3 is negative, meaning that increases in test scores are associated with a lower probability of leaving a particular school. The identifying assumption is that controlling for $score_{st}$ absorbs any pre-existing relationship that exists between test scores and school or neighbourhood characteristics, and that any change in this relationship is exogenous because the only thing that changes over this relatively-short time period is that the information has been revealed. School and student characteristics are added to further absorb any residual relationship that may exist between test scores and school/neighbourhood demographics.

One concern is that this analysis will not be able to separate a poor reaction to test scores from a constrained reaction due to high switching costs. At the time the test scores were revealed, students were required to attend their neighbourhood area school (catchment area school), and thus it was relatively costly and difficult to switch schools because it meant that a family needed to switch residences. Before describing how I address this problem in the analysis

[9]Test scores are averaged together to make the analysis simpler, and because the scores are highly correlated, making it difficult to estimate their independent effects in the regression framework.

note that in Table 3.3, in spite of potentially high switching costs, between kindergarten and grade 1 about 10 to 13 percent of all students switch schools. Furthermore, nearly 30% of students in the sample switch schools at least once between kindergarten and grade 5. Thus, even in the presence of high costs, there are a substantial number of students switching schools.

More formally, I use two strategies to show that switching costs are not the primary driver of the results. First, in the context of Equation 3.9 I look at how α_3 changes after BC implemented its open boundaries legislation. If costs are driving the results, then this coefficient should be much stronger once the legislation is enacted.

The second strategy relies on the assumption that costs should be less of an issue for parents' initial school choice. In this approach, I aggregate the data to the school level look first at kindergarten enrollment changes as a function of school test scores. If costs are less of an issue for the initial choice of school, I should observe larger enrollment increases in schools that score better on the test. The estimating equation will be

$$\frac{enr_{st+1} - enr_{st}}{enr_{st}} = \alpha_0 + \alpha_1 score_{st} + \alpha_2 d^{after} + \alpha_3 score_{st} \times d^{after} + X_{st}'\beta + \delta_s + \epsilon_{st} \tag{3.10}$$

where enr_{st} is kindergarten enrollment in school s at time t, and all other variables are as previously defined. Percentage change in enrollment is used as the dependent variable because of the wide variance in school size. α_3 is still the coefficient of interest, representing the increase in percentage change for schools scoring higher on the standardized test.

While aggregate analysis does generate interesting results in terms of total movement, it cannot answer any questions about individual heterogeneity in the effects. Even though an individual switching analysis cannot be performed for kindergarten choice because students are not observed before attending school for the first time, it remains possible to study hypotheses at the individual level within a discrete choice analysis framework. I use the Conditional Logit from McFadden (1974), which models the choice of school among a set of alternatives as a function of school attributes. In this model the probability that student i chooses school s is

$$P_{is} = \frac{e^{X_s'\beta}}{\sum_{\in k} e^{X_k'\beta}} \tag{3.11}$$

where X_s' is the full set of variables thought to influence school choice, including test scores. Estimating the relationship between test scores and choice in this way allows for a comparison of preference weights placed on different characteristics of the school, such as neighbourhood attributes, travel distance, and school test scores. It also allows the preference for school test scores to depend on individual characteristics, so that I am able to compare the different weights placed on test scores for varying segments of the population.

3.5 Data

3.5.1 Description

The main data comprises four pieces. The first piece is a set of administrative records collected by the Ministry of Education in BC on all students in the public and private school system from 1991-2004.[10] Students are tracked through time by a personal identification number, and are observed in the data as long as they are registered in the BC school system. Most importantly for this analysis, the data contain a school number for each student in each year, so it is known when a student switches schools. The data are geocoded for all records after 1995 at the 6-digit postal code level, which is an area as small as one side of the street on a city block in urban locations. Because the area is so small, postal code changes are used to determine residential changes, which will be very accurate unless people frequently switch to a residence in their immediate vicinity. Finally, the data contain a small set of personal characteristics including gender, aboriginal status, home language, English as a second language (ESL) status, special education (SpEd) status.

The second piece is information collected by the BC Ministry of Education on all public and private *schools* from 1991-2004. Schools are tracked over time by an id number, which is important because a small number of schools switch physical locations. Information is given on whether the school is standard or alternative, whether it is public or private, its open and close dates, grades offered, teacher counts, the name of the principal and exact address and postal

[10]Private schools are called *independent* schools in BC, but for convenience I will continue to call them private schools.

code. The school data are geocoded using the exact address where possible, making it possible to calculate distance to school for each student.

The third piece is a collection of test scores from 1999-2003 on the Foundation Skills Assessment (FSA). The data contain individual-level raw scores on the numeracy, reading, and writing tests for grade 4, 7, and 10 students. All students who are eligible to write the test are contained in the data, whether or not they actually write it. If a student is excused from the exam they are flagged in the data, though no reasons are given for their absence.

The final piece is Canada Census data from 1996 and 2001 at the Enumeration/Dissemination Area (EA/DA) and Forward Sortation Area levels.[11] This data is brought into the analysis to proxy for parent income, education, and demographic information not contained within the BC data obtained from the Ministry. The EA/DA data is attached to student records to proxy for student-level income and education, while the Forward Sortation Area level data is attached to the school data to proxy for the characteristics of the neighbourhood that surround the school. While Forward Sortation Area data are attached directly to the schools data, the procedure by which the EA/DA level data is slightly more complicated, and it is described in the Appendix.

Though the data are well-suited for this analysis, there are some caveats that are important to keep in mind. First, data are collected once per year, so if a student switches schools, it is not known at what point in the school

[11]DA is the name given in the 2001 Census to a relatively stable area targeted to contain 400-700 people. EAs are similarly defined areas from the 1996 Census. Though the boundaries for these areas are different, they are the best proxy available for household income that can be attached to this data. The Forward Sortation Area is the first three digits of the postal code, and roughly represents a large neighbourhood.

year the switch occurs. Second, no reasons are given for school switches, so we cannot decipher if a switch is due to school quality, job changes, or one of the many other reasons children switch schools. Nevertheless, the ability to track students from year to year and observe their switches should provide enough information on mobility to reveal effects of information shocks, should they exist.

The analysis focuses on the mobility of elementary school students, since they are arguably the most interesting subpopulation for such a study. The future outcomes of younger children are affected by decisions made in elementary school, so parents will likely be very concerned with school quality at that stage. If there is to be a reaction to new information on quality, it is likely to be among elementary school parents. Furthermore, there are many more elementary schools than secondary schools, so the amount of choice a parent faces is much higher, meaning a greater chance to observe school switches.

I define the sample universe as students who live in the Lower Mainland, are in grade 5 and below, whose initial entry into school is observed in 1992 or after, and who attend schools that remain open and offer all grades over the sample period. The Lower Mainland comprises the areas surrounding Vancouver, and the only populous area in BC, so it is natural to restrict the analysis to this area. The reason for the latter restrictions is to make sure that observed movement between schools is voluntary, and not the result of forced moves from school closures or redrawn school catchment areas resulting from school openings. In total there are 1,112,609 observations on 239,315 students.

From this universe, several exclusions are made, all of which are quantified in Table 3.1. First, students who are not observed in every year since entry (8.10%), are dropped. The reason is that no data are collected for students who exit the system, so we would not know anything about their mobility. Second, students who skip or repeat grades (1.29%) are excluded because they represent a small, special sample of kids whose movement I am not trying to capture. Finally, students who attend a school outside of the Lower Mainland at some point between kindergarten and grade 5 (4.15%) because they pose difficulty for tracking. In practice, these exclusions amount to very little of the sample universe, leaving about 86% for analysis.

3.5.2 Variable Definitions, Timing, and Unobserved Test Scores

It is important to be clear on the definition of switching I am using in the empirical analysis. In all regressions, switches are defined as school changes between years t and $t+1$. Defining switching in this way means that I am looking exclusively at the effect of test scores on switching *out* of your current school. Alternatively, I could have examined switches between $t-1$ and t, which is interpreted as switches *into* the current school. While both dependent variables are interesting in their own right, I chose to focus on school exits because this behaviour could be driven more by test score results than school entries given constraints that exist in BC. For example, a group of parents who leave a school scoring particularly poorly on the FSA test may want to enroll their child in a school nearby that scored well on the test. Due to capacity constraints, all such students might not be admitted to the better school, and might be distributed to other schools based on available space. In this situation, while a strong relationship would be observed when looking at school exits, a potentially much weaker relationship would emerge using school entries as the dependent variable.

A follow-up concern is what test score information is observable to parents making a switch between t and $t+1$? To determine the answer to this question, Figure 2 depicts the timing of the student data collection and the release of the first round of test scores. The structure of the data is such that information on students is collected once per year just after the beginning of the school year in October. The FSA test is administered once per year at the end of the school year in April or May, and the results are released in November of the following school year. The important thing to note is that the scores from the test written in year $t-1$ are released *after* the data collection date in year t. Thus, the information available to any individual deciding to switch schools between t and $t+1$ is the test score from $t-1$. Empirically, I attach the first lag of test scores to all schools.

The data contain individual-level test scores for grades 4, 7, and 10 in the areas of numeracy, reading, and writing. Because the analysis focuses on elementary students, I use only the grade 4 test, which would be most relevant to this subgroup. Furthermore, I use only the numeracy and reading test information, as it is not clear how much useful information is contained within the writing scores. To arrive at a score for each school, I first standardize the numeracy and reading tests by year to have a mean of 0 and standard deviation of 10 at the student level. Then I take the average score on each test for each school, and combine the numeracy and reading scores by averaging them together with equal weight. Thus, for each school in each year, there is exactly one test score.

A final issue is that no school has any test score observations before school year 1999 (the first year the scores were available). This poses a problem because the analysis focuses on comparing switching rates before and after the information was released. The solution I adopt in this paper is to predict school test scores for 1998 based on the five years of observable school-level test score data using the following specification:

$$score_{st} = \alpha_0 + \alpha_1 year_t + \delta_s + \delta_s \times year_t + v_{st} \tag{3.12}$$

where $year_t$ is a time trend and δ_s are school fixed effects. This specification essentially estimates a separate linear time trend for each school, which is then used to predict the value of score in year 1998. Given that this specification estimates the trends in test scores by school with only 5 observations, this method has limits. It is, however, the best available method for estimating test scores for a year when they are not observed.[12]

[12] I have also run regressions with a quadratic trend and obtained similar results throughout. A cubic trend was also attempted, but this tended to produce fairly wild results, so that method was abandoned. Finally, I also attempted to simply assign the year 1999 test score to schools in year 1998, and this method also yielded fairly similar results.

3.5.3 Summary Statistics

Table 3.2 contrasts means of variables used in the analysis by whether the student has ever switched schools voluntarily. Though switchers and non-switchers are clearly different on a number of dimensions, they are actually quite similar overall. The notable differences are a higher fraction of switchers do not speak English at home, have attended an ESL program, and are observed in public schools only. Furthermore, the household incomes in the areas in which they live are about $5,000 less than the non-switchers. Most other characteristics are quite similar, with small differences on some characteristics suggesting that switchers come from different backgrounds. Note, however, that both switchers and non-switchers have about the same number of public and private schools within 5 km of their residence, so access to alternative schools does not vary across groups.

Table 3.3 presents the fraction of students who switch by grade across school entry cohorts, and separates them by income level. These fractions are separated by entry cohort to examine interesting time variation, and by income to see if one particular group is driving the results. The first two columns in Panel A describe the fraction of students that have never switched and switched once. The vast majority of students never switch schools voluntarily, and most of those who do switch do it one time. These fractions appear to be relatively stable, but show a slight tendency towards less switching as time goes on. The remaining columns in Panel A calculate the fraction of students in grade g who switch as they move into grade $g + 1$. There are many more moves between kindergarten and grade 1 than at any other point in the school career, about 13% in the early years decreasing to less than 10% in the later years. The fraction declines into the higher grades to about 7% switching between grades 4 and 5.

Panels B and C check for differences in the propensity to switch by income level. Comparing panels, students living in high income areas are much less likely to switch in their lifetime, and a smaller fraction of them are switching at each and every stage in school. Thus, the estimates presented later in the paper are driven mostly by the behaviour of lower income people, even though a non-trivial fraction of high income individuals are mobile.

Table 3.4 looks more closely at the destinations of switchers broken down by school year. The first two columns take the set of all switchers and calculate the fraction that switch postal codes and switch district. Because of the restrictions in BC requiring students to attend their catchment area school, the majority of school switches are accompanied by residential moves. Note, however, that the neighbourhood school restriction was lifted by legislation introduced in 2002, which explains the lower fraction of joint school and residence switchers in the last couple of years in Table 3.4. About one-third of all switchers also change districts, representing possibly their distaste with the entire school board and not just the school.

The remaining columns present selected statistics separated by district switching status. I look at whether students who switch districts are more likely to move to a school located in a neighbourhood with higher income and a higher fraction of university-educated people. Just shy of 60% of district movers locate in areas with higher incomes, versus about 35% of non-district movers. District movers are also more likely to pick areas with a higher fraction of university-educated people, at 45% versus 35% for non-district movers. Based on these statistics, it appears as though district movers could be more likely to pick better schools than district stayers.

At the school level, Table 3.5 compares characteristics schools that scored above and below the median on the FSA test. As expected, schools that score well on the test have favourable demographic and neighbourhood characteristics. Most notably, for above-median schools the average household income level is about \$10,000 higher, average dwelling value is about \$63,000 higher, there is a higher fraction of university-educated individuals, more students speaking English at home, and a lower fraction are public schools. The results in Table 3.5 are important because they highlight the effects of residential sorting, and the need to control for these confounding factors in the empirical analysis.

3.6 Analysis

3.6.1 The Effect of Test Scores on School Switching

This section presents the regression analysis of the effect of test scores on the probability that a student switches schools between years t and $t+1$. The Tables below present only the main coefficients; the full results including all coefficients from control variables are available upon request. The estimates are obtained from a Linear Probability Model (LPM) because the preferred specification contains fixed effects, which can cause problems with consistency in non-linear models.[13] Models are estimated only for public school students, since a separate analysis of private school students revealed no interesting estimates. The results that follow show that there are small decreases in the probability of switching out of your school in response to higher school-level test scores. Further, it appears that this effect gets somewhat stronger over time, but does not vary much across student characteristics.

Table 3.6 reports the *initial impact* of increased test scores on the probability of switching. In these regressions, only data from years 1999 and 2000 are included, where 1999 is the year before the scores were available and year 2000 is the first time parents could respond after seeing the information.[14] In this Table and those that follow, the dependent variable has been rescaled to be 0 or 100 rather than 0 or 1 so that the coefficients represent percentages rather than decimals. Note also that because the independent variable has been scaled to have a standard deviation of 10, the coefficients represent the impact of a $1/10\sigma$ change. In practice, this amounts to about 2-3 percentage points on the FSA test. The coefficient of interest is the interaction between $score$ and d^{2000}.

Column 1 is a baseline estimate with no control variables. The coefficient on the interaction term shows that increasing the test score by $1/10\sigma$ leads to a 0.180 percentage point decline in the probability of switching schools. Comparing this effect size to the mean of the dependent variable (7.9%), it can be characterized as small. The main effect of $score$ in these regressions

[13]In spite of this, estimates from baseline regressions using a logit are very similar to the LPM estimates, and are available upon request.

[14]Because test scores are lagged by 1 year, 1999 is the first year they were available, and not 1998.

is over twice the size as the interaction, and statistically significant, meaning that even in 1999 before the scores were observed, there is less switching out of schools that score higher on the test.

Column 2 adds control variables at the school level, leading to a decline in the interaction effect to -0.134. Interestingly, school controls reduce the main effect of *score*, as they directly control for some of the correlation between test scores and neighbourhood characteristics. Column 3 adds student level controls, reducing the interaction effect slightly to -0.121 and the main effect to -0.058 and statistically insignificant. Finally, the preferred specification in column 4 adds a set of school fixed effects to control for time-invariant characteristics of schools, which settles the interaction term to -0.100, which is statistically significant at the 10% level, and just barely insignificant at the 5% level. As a robustness check, Appendix Table 3.B1 replaces the average test score with the .75 quantile and arrives at quantitatively similar results.

In the preferred specification, the impact of increasing the test score by $1/10\sigma$ leads to a 0.100 decline in the probability of switching. The independent variable was scaled in this way so that the test score changes represented a "realistic" difference in test scores between two schools. A more extreme change of 1σ would be like comparing one of the worst schools to one of the best schools, and in this comparison the difference in probability of switching is over 1 percentage point, or about 14% off the mean of the dependent variable. From this perspective, increases in test scores do lead to an economically significant difference in the probability of switching out of a school, even though smaller differences between schools in terms of test scores do not seem to have a large effect.

Table 3.7 examines what happens as more test scores are revealed over time using the fixed effects specification. The regressions in this Table contain data from 1999 - 2003, and has two sets of columns. In the first two, the test score for school X in year t is used as the independent variable. For fear that this may be too noisy a measure, columns 3 and 4 use a two-year moving average of test scores. For example, the score for school X in year 2002 is the average of the scores obtained in 2001, and 2002. The year dummies in these regressions are defined as $d^t = 1\{year >= t\}$ rather than $d^t = 1\{year = t\}$ so that their coefficients represent the *increment* between year $t-1$ and t in the coefficients. They are defined in this way to focus attention on how the effect changes from year to year.

In column 1 the data are pooled and I estimate the cumulative effect of increasing test scores on switching in the years after scores are observed. The estimate of the interaction effect implies a 0.120 percentage point decrease for each $1/10\sigma$ increase in test scores. This estimate is more negative than the one obtained in Table 3.6, lending support to the conclusion that parents place more weight on these scores as more of them are observed. Column 2 breaks the effects down by year to see how the effect evolves over time. The fact that all the interactions are negative implies that the effect gets more and more negative as time passes, though the change in the magnitude cannot be statistically distinguished from zero except in the initial year. The estimates from columns 3 and 4 using the smoothed test scores provide slightly stronger results suggesting that perhaps parents do react to an average of previous information rather than just the observation in a particular year.

Of particular importance is the coefficient on $score \times d^{2002}$, since this is the first year that legislation was introduced allowing parents to attend any school in the province. Using both raw and smoothed test scores, this coefficient is larger than the surrounding years, and statistically significant at the 10% level in the case of the smoothed scores. Given the stronger response, it appears that constraints are playing a small role in limiting the mobility of parents, and once these constraints are lifted there is more scope to react to the results of the test. Even so, the coefficients are not extremely large, meaning that the binding constraint to mobility is moving costs and not the ability to gain admission to a school outside one's catchment area.

Table 3.8 takes data from 1999 - 2003 and interacts $score \times d^{2000}$ with a set of personal characteristics of students to check whether there is heterogeneity in the response to test scores. A priori it is not clear exactly who might have a greater response to the scores. For example, though previous research has shown that lower-income individuals care less about academics when choosing a school, they might make greater use of the test score if they otherwise cannot afford the time to visit schools and obtain quality information by other means. As another example, recent immigrants might make greater use of the test scores if they do not have a good understanding of the school system, though they may not if they do not know how to interpret them. Thus, exactly who reacts to the test scores by more is an empirical question. Note that the variables have been scaled such that income represents a \$1000 change and % dropout represents a 10-percentage point change. The values of all variables in the interaction terms have been demeaned, so that the main effect of test scores is for the average individual.

The general conclusion of Table 3.8 is that in this data there is very little heterogeneity in the response to the test scores. In column 1, $score \times d^{2000}$ is interacted with measures of the amount and quality of choice in nearby schools. Interacting with the number of public and private schools within 5 km of the current school does not lead to appreciable changes in the coefficient, though it is interesting to note that when more private schools are located nearby, higher test scores imply less of a decrease in the probability of switching out. Conversely, the less private schools are located nearby, the lower the probability of switching in response to good information about your school. This result makes sense given that with less choice, a parent should be more inclined to stay in a better-performing school because switching costs are much higher. Parent reactions do not vary with the number of public schools located nearby, but do vary with the quality of those schools. The decrease in the likelihood of leaving a school in response to good information is stronger when the quality of schools nearby is lower by 0.023 percentage points for every $1/10\sigma$ decline in the nearby-school average. Thus, students located in areas with very poor alternatives are likely to stay in their current school if they receive good information.

Column 2 looks at the interaction with neighbourhood income and education characteristics. Though none of the estimates are statistically significant, the decrease in the likelihood of leaving a school in response to good information is stronger when education in the area is lower. This could mean that lower-income parents are reacting more strongly to the test scores than others, which would make sense if they had weaker prior information on school quality. Income appears to have no bearing on the magnitude of the test score reaction.

Column 3 checks for nonlinearities in the test score effect by checking whether the square of the test score changes after test scores are observed. The positive coefficient means that the decline in the probability of leaving a school in response to good information is even stronger when the school has a bad score. One interpretation of this result is that a good test score from a bad school is more of a positive signal than a good test score from a good school, and as such parents will want to stay put.

Finally, column 4 interacts the test score with a measure of student ability and an indicator for whether the student speaks English at home. Ability is the student's score on the grade 4 test, which I attach to all observations on that student. Thus, I treat a student's grade 4 test score as a measure of their natural ability for all years they are observed in the data. The sample size is smaller in this exercise because some students are never observed writing the test. The reaction to the test score information is only very weakly related to student ability, though it is much stronger for students who do not speak English at home. The negative reaction to good test score information is weaker if the student is a non-English speaker, meaning perhaps that non-English speakers either have a harder time interpreting the information or that they simply do not use test scores as much as English-speakers.

3.6.2 The Effect of Test Scores on Kindergarten Enrollment and Choice

The small size of the coefficients in the switching regressions might be due in part to high switching costs preventing parents from acting on the new information they receive from test scores. This section tests this hypothesis by examining the effects of test scores on initial school choice. Costs might be less of a concern to parents in making their initial school choice because they anticipate having to incur more costs in the future if they make a bad choice now, and also because they want to make their child's first schooling experience as positive as possible. To see whether reactions to test scores are stronger for the initial school choice, I conduct two separate analyses. The first looks at whether enrollment changes from year to year are different for good and bad schools after test scores are revealed. While this is informative, the downside to an aggregate analysis such as this is that heterogeneity in the response at the individual level cannot be examined. For this reason, the second analysis presents results from a Conditional Logit model of school choice at the individual level.

Table 3.9 reports coefficients from a regression of percentage change in enrollment on school test scores over time using school fixed effects. This Table is structured in the same way as Table 3.7, meaning that it contains regressions using both the raw and averaged test scores as covariates. Like the previous regressions, the coefficients of interest are the interactions between *score* and d^t. Column 1 presents the cumulative effect of test scores on percentage change enrollment in kindergarten pooling all years 1999 to 2003. There is a very modest positive effect: a $1/10\ \sigma$ increase in test scores increases the percentage change in enrollment by 0.5 percentage points. Compared to the mean percentage change enrollment in the sample of 7.8%, these are not extremely large effects, though they appear to be slightly stronger than the effect of test scores on switching.

Column 2 breaks the change down by year. The coefficients are in general small and statistically insignificant and exhibit a positive to negative switching pattern in alternating years. The two positive effects are the initial impact in year 2000 and the effect in year 2002. The year 2002 effect is the strongest and only statistically significant result because of the open boundaries legislation introduced that year, supporting yet again the notion that enrollment restrictions matter when reacting to test score information. The 2002 effect implies that increasing test scores by $1/10\sigma$ increases the percentage change in enrollment by 1.7 percentage points, or 2.1 percentage points using a smoothed test score, an effect that is non-trivial in relation to the mean percentage change in enrollment.

It is interesting to note the changes in the size of the intercepts. Most notably, between 2002 and 2003 the percentage change in enrollment jumps up by about 5 percentage points. Thus, not only does the open boundaries legislation allow parents to react more to the test score information, it permits a large increase in mobility overall.

Table 3.10 presents the results of the discrete choice analysis at the individual level. For these regressions, the full sample of data was too large to be able to estimate the coefficients. As a result, the sample was restricted to all schools in the cities of Vancouver and Burnaby, since they are two very populous cities that are not separated by large rivers like North Vancouver and Surrey (see the map in Figure 3.7). The signs of the coefficients determine the sign of the effect of changing attribute x of school s on the probability of choosing school s, and the negative of the sign of the effect of changing attribute x of school s on the probability of choosing school k. The magnitude of the coefficients do not have any particularly interesting meaning except when compared in size to other coefficients. Recall that the effect of test scores is measured for a $1/10\sigma$ change, income represents a \$1000 change, the neighbourhood attributes represent a 10-percentage point change and distance is in kilometers. The values of all variables in the interaction terms have been demeaned, so that the main effect of test scores is for the average individual.

From Table 3.10 we see that post-2001 a higher test score for school s leads to a higher probability of choosing a school s and a lower probability of choosing any other school k.[15] Comparing magnitudes, a 1σ change in test scores is equivalent to an 8.76 percentage point decrease in the proportion Aboriginal, 56.52 percentage point decline in the unemployment rate, a \$12,380 increase in the income of the neighbourhood around the school, and a 0.16 km decrease in distance from school. Clearly, the strongest preference is for distance to school, which may be driven in part by restrictions on attending the neighbourhood school.

As indicated by the three-way interaction terms, this preference for test

[15]Here I use post-2001 because I am estimating the probability of choosing a school in yaer t, whereas previously I looked at the probability of switching between t and $t+1$. For this exercise, the test score has to be lagged twice.

scores does not appear to vary with income, but does depend negatively on the percent dropout in your neighbourhood and on whether you speak English at home. For an individual with average income who speaks English at home, a 20 percentage point increase in the fraction of dropouts in the neighbourhood would be enough to make the effect of test scores equal to zero. The same is true for a switch to not speaking English at home for an individual with average income and an average fraction of dropouts. Thus, for combinations of non-English speakers living in poorly educated neighbourhoods the effect of test scores is actually negative. This is exactly the type of group that might face challenges in interpreting the test score information, or may simply care less about academics when picking a school for their child. Note that these results are at odds with the results from the previous section that showed people living in lower educated areas and who do not speak English at home reacted more strongly to test scores. This could mean that this group of people do not make good use of this information in their initial decision, but perhaps understand more its importance when choosing to switch schools.

3.6.3 Discussion

The main goal of the empirical analysis was to check for real-world support for the theoretical predictions that parents use new information when it becomes available, shift more weight to the new information when they observe several pieces, and have heterogeneous reactions. The analysis of the BC data supports the first two predictions by showing that there is a small decrease in the probability of switching out of your school when you receive good test scores, and that this decline gets stronger when more test scores are observed. Furthermore, kindergarten enrollment increases in schools that score well on the test, though one caveat is that it does not necessarily get stronger over time. The data do not strongly support the heterogeneity prediction, though there is some evidence that the response varies according to the amount and quality of choice available nearby, education levels in the residential neighbourhoods, and home language spoken.

The responses to the test scores are small, which is consistent with the literature examining preferences for school quality. It is possible that small response size may be due in part to one of a few types of costs facing parents as they decide to switch schools. The first type is administrative: parents are simply not allowed to attend schools other than their neighbourhood school. The estimates do support some role for these costs, as the coefficients on test scores in both the switching and kindergarten enrollment regressions increase in 2002 when BC dropped its neighbourhood school boundaries requirement. It is important to keep in mind, however, that the increase is not particularly large, suggesting a fairly small role for those types of costs.

The second type is a moving costs: many school moves also require residential moves, and this makes it very costly to switch schools. Indeed, it is likely true that more than 8 - 12% of children would switch schools every year if moving costs were nil, and this is true whether or not a school choice program exists. The question is whether reducing moving costs would lead to a larger response. I examine this issue in two ways. First, in Table 3.7 I interacted test scores with a measure of the number of public schools within 5 km, the hypothesis being that with more choice nearby, parents are less likely to have to switch residences or incur large moving costs. The results is that the main effect does not vary with the amount of local choice. The second way was to look exclusively at test scores' effect on kindergarten enrollment under the assumption that moving costs were less important for initial school choices. The estimates were not overwhelmingly large, suggesting again that moving costs may not be playing a large role. The combined evidence on moving costs suggests that while it is likely to play some role in keeping effect sizes small, it is probably not a large role.

A final type of cost is mental or emotional. It may be difficult for a child to leave the comfort of their school and friends to move to an entirely new environment. If parents take this into consideration, they may not move a child even after observing that its test scores are poor.[16] While I cannot directly address this issue in a regression, some evidence against this hypothesis is that there are many children that move every year, up to 12% from kindergarten to grade 1. Even if there are some parents that would be unwilling to move because of mental costs, there is still a large fraction of parents willing to switch their child to a different school, and it should be enough to identify an effect.

Perhaps the low response is driven by the inability of parents to easily access this information, and if parents had an easier way to get ahold of test score information, then effect sizes would be stronger. If an experiment like

[16]It could also be the case that students leave a school where they are facing difficulties with teachers or peers, and the move would represent a mental or emotional benefit. In this case I would not expect any effect on the estimates.

Hastings and Weinstein (2007) were conducted then the effect sizes would certainly be stronger, as they show in their study. Yet, even in the absence of a single sheet of paper with containing average test scores, the BC Ministry of education does provide an easy tool to get information on individual school test score results. This website provides details about changes over time, and compares school results to the district and the province in an easy to interpret manner. Furthermore, the Fraser Institute, a Canadian think tank located in BC, publishes school report cards annually that rank schools based on the FSA test scores. Thus, parents are very likely to have seen the test score information, and cannot be fully explained by an inability to access it.

If none of these factors is the driving force behind the small estimates, the remaining explanation is that they simply do not place a large weight on these scores. The model contains several explanations for why this might be the case. First, it could be that test scores simply do not contain enough new information to be that useful to parents. This is definitely possible, especially given results from Johnson (2005) that neighbourhood and school characteristics easily available in the Census can explain between 40 and 50% of the variation in school test scores. Another possibility is that the test scores were simply perceived as being too noisy, and therefore parents did not place much weight on them. This too is possible, given that using standardized testing to evaluate a school is heavily criticized by teacher unions and other education professionals. Furthermore, this particular testing system is new, and it may take time before people perceive that information to be useful. A final possibility is that prior information was very good on school quality, and so parents put relatively little weight on the test as a result. This is the least likely of the three explanations, given how inherently difficult it is to obtain precise school quality information without incurring a large cost.

There has been a strong movement towards outcomes-based school accountability in North America, particularly in the United States with the No Child Left Behind Act of 2002. Part of the aim of programs such as NCLB and the FSA test in BC is to publish the results of standardized tests to parents so that they have more complete information, and can therefore create competitive pressure for schools to produce education or risk losing enrollment. This demand response by parents is important because without it schools would have reduced incentives to improve. Based on the results of this paper, it is unlikely that adding test score information has altered the incentives that schools face substantially, and it is questionable whether this benefit added by publishing these scores is high enough to outweigh the cost.

3.7 Conclusion

In this paper I model the effect that new information has on decisions regarding school choice, then test the predictions of the model using a natural experiment that increased parents' information sets. The general conclusion is that parents react to good test score information about their school by not switching out of that school. Furthermore, using a sample of kindergartners I show that enrollment flows into schools that score well on a standardized test, and supplementary evidence at the individual level appears to back up those findings. Overall, the magnitude of the response is small, which may be driven in small part by barriers to mobility such as costs and administrative rules. I support this statement by showing that the response is only a bit larger when BC government implemented its open boundaries legislation, allowing children to attend any school in the province room permitting. There is not a great deal of heterogeneity in the response, though it is clear that whether an individual speaks English at home affects how parents react to test scores.

The small response implies that there may not be much extra competition generated by increased mobility of parents, and as such the potential educational efficiency increases may not be realized using such an intervention. While test scores are clearly useful as a summary measure of student performance, parents seem to be well-informed already about the information contained within those test scores. If parents already have a good idea about the quality of schools, the benefits of publishing such test scores may not outweigh the costs. Instead, these funds could be channeled in such a way as to lower the cost of obtaining information, which would likely have a large impact on mobility if the results of other research can be generalized to the BC setting.

References

Altonji, Joseph G., and Charles R. Pierret (2001) 'Employer learning and statistical discrimination.' *The Quarterly Journal of Economics*

Barrow, Lisa (2002) 'School choice through relocation: evidence from the washington, d.c. area.' *Journal of Public Economics* 86(2), 155–189

Bayer, Patrick, Fernando Ferreira, and Robert McMillan (2004) 'Tiebout sorting, social multipliers and the demand for school quality.' NBER Working Papers 10871, National Bureau of Economic Research, Inc, November

Beaulieu, Nancy Dean (2002) 'Quality information and consumer health plan choices.' *Journal of Health Economics* 21(1), 43–63

Black, Sandra E. (1999) 'Do better schools matter? parental valuation of elementary education.' *The Quarterly Journal of Economics* 114(2), 577–599

Chernew, Michael, Gautam Gowrisankaran, and Dennis P. Scanlon (2001) 'Learning and the value of information: Evidence from health plan report cards.' NBER Working Papers 8589, National Bureau of Economic Research, Inc, November

Erdem, Tulin, and Michael P. Keane (1996) 'Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets.' *Marketing Science* 15(1), 1–20

Figlio, David N., and Maurice E. Lucas (2004) 'What's in a grade? school report cards and the housing market.' *American Economic Review* 94(3), 591–604

Gibbons, Steve, and Stephen Machin (2003) 'Valuing english primary schools.' *Journal of Urban Economics* 53(2), 197–219

Hastings, Justine S., and Jeffrey M. Weinstein (2007) 'No child left behind: Estimating the impact on choices and student outcomes.' NBER Working Papers 13009, National Bureau of Economic Research, Inc, April

Hastings, Justine S., Richard Van Weelden, and Jeffrey Weinstein (2007) 'Preferences, information, and parental choice behavior in public school choice.' NBER Working Papers 12995, National Bureau of Economic Research, Inc, March

Hastings, Justine S., Thomas J. Kane, and Douglas O. Staiger (2005) 'Parental preferences and school competition: Evidence from a public school choice program.' NBER Working Papers 11805, National Bureau of Economic Research, Inc, November

Hussain, Iftikhar (2007) 'School choice, information disclosure and sanctions: Evidence from an unusually tough school accountability regime.' Mimeo, University College London

Ichino, Andrea, and Enrico Moretti (2006) 'Biological gender differences, absenteeism and the earning gap.' NBER Working Papers 12369, National Bureau of Economic Research, Inc, July

Jin, Ginger Zhe, and Alan T. Sorensen (2006) 'Information and consumer choice: The value of publicized health plan ratings.' *Journal of Health Economics* 25(2), 248–275

Johnson, David (2005) *Signposts of Success - Interpreting Ontario's Elementary School Test Scores* (Toronto: C.D. Howe Institute)

Lange, Fabian (2007) 'The speed of employer learning.' *The Journal of Labor Economics* 25(1), 1–34

McFadden, Daniel (1974) 'Conditional logit analysis of qualitative choice behavior.' In *Frontiers in Econometrics,* ed. P. Zarembka (New York: Academic Press) pp. 105–142

Moretti, Enrico (2008) 'Social learning and peer effects in consumption: Evidence from movie sales.' Mimeo, University of California Berkeley

Pope, Devin G. (2006) 'Reacting to rankings: Evidence from "america's best hospitals and colleges".' Mimeo, University of California, Berkeley

Scanlon, Dennis P., Michael Chernew, Catherine McLaughlin, and Gary Solon (2002) 'The impact of health plan report cards on managed care enrollment.' *Journal of Health Economics* 21(1), 19–41

Wedig, Gerard J., and Ming Tai-Seale (2002) 'The effect of report cards on consumer choice in the health insurance market.' *Journal of Health Economics* 21(6), 1031–1048

Table 3.1: Sample Exclusions

	Students	% of Total	Observations	% of Total
Universe	239,315	100	1,112,609	100
Exclude:				
Not observed all years	19,384	8.1	55,459	5.0
Repeaters/Skippers	3,094	1.3	18,570	1.7
Outside Urban Area	9,936	4.2	5,200	5.0
Final Sample	206,901	86.5	983,380	88.4

Universe consists of K-5 students, entering 1992+, observed at least once in Lower Mainland, attending schools offering K-5 observed over all years, who we can rank with test scores. Fractions in columns 2 and 4 differ slightly when the order of exclusions is altered, because of shared characteristics. The sample period runs from entry cohort in 1992 - 2004, and contains observations through time for grades kindergarten through 5. The final sample is a balanced panel at the student level.

Table 3.2: Means of Personal Characteristics by Switching Status

	All	Switchers	Non-Switchers
Male	50.9	51.2	50.8
Aboriginal	4.0	6.2	3.3
Non-English	30.8	37.7	28.5
Ever ESL	28.1	35.9	25.5
Ever Special Ed	6.6	9.3	5.7
Ever Private School	11.4	11.5	11.4
Household Income	65540.4	61517.9	66840.2
S.E. of Household Income	6358.8	5724.9	6563.7
%Education < Gr 9	7.2	7.8	7.0
% Education University	27.8	26.2	28.4
% Immigrant	33.7	34.1	33.5
% Visible Minority	32.9	34.0	32.6
% 1-Parent Families	14.9	15.9	14.6
% 25+ Unemployed	6.5	7.2	6.2
Avg Dwelling Value	290125.3	275692.2	294789.2
% Dwellings Owned	70.6	67.8	71.5
% Mobile Last Year	15.8	17.2	15.3
% Mobile Last 5 Years	47.6	50.5	46.7
Travel Distance to School	1.7	1.9	1.7
# Public Schools < 5km	2424.3	2517.7	2394.1
# Private Schools < 5km	3.7	3.7	3.6

Household income, the standard error of household income, and average dwelling value are all in year 2000 dollars. All other variables except the final two are percentages. Switching and all other personal characteristics refers to whether the person ever possessed that characteristic. Continuous variables are averaged over the number of years the student is observed. Income, education, employment and immigrant variables are linked to student records from the census, and represent the percentage of people in that student's neighbourhood with those characteristics.

Table 3.3: Propensity to Switch by Grade, Income Level and School Entry Year

	Never Switch	Switch Once	Switch K-1	Switch 1-2	Switch 2-3	Switch 3-4	Switch 4-5
Panel A: All Students							
1992	68.1	23.3	12.9	9.3	7.4	7.1	6.7
1993	68.8	22.7	12.4	7.8	7.7	7.9	6.5
1994	70.0	22.2	10.6	8.6	8.1	7.0	6.3
1995	69.7	22.1	11.5	9.0	7.6	6.7	6.3
1996	69.8	21.9	12.3	8.2	7.5	7.2	5.9
1997	70.5	21.6	10.6	8.2	6.9	7.1	7.0
1998	71.1	21.7	10.3	7.4	6.4	7.2	6.9
1999	70.2	22.4	9.7	7.4	7.7	7.5	7.3
Panel B: Low Income							
1996	63.4	25.9	14.6	10.4	9.5	9.3	8.1
1997	65.3	25.0	12.3	10.0	8.9	9.2	8.8
1998	65.9	24.9	12.6	9.0	7.9	9.0	8.6
1999	63.5	26.7	12.1	9.6	9.9	9.5	9.0
Panel C: High Income							
1996	78.1	16.6	9.3	5.5	4.9	5.4	4.2
1997	77.9	16.8	8.2	5.6	5.1	5.2	5.4
1998	78.4	17.2	7.1	5.9	5.0	5.6	5.4
1999	76.7	18.0	7.4	5.3	5.7	5.7	5.8

All values in the table are percentages. Entry cohort is the first year the student enters kindergarten. Never switch means that the student remains in the same school between kindergarten and grade 5. Switch K-1 means the student switched schools between kindergarten and grade 1. Low income is defined as being below the median income level, and vice versa for high income. Low and high income panels have less entry cohorts because the first year census data can be linked to the student data is in 1996.

Table 3.4: Description of Switching Type

	All Students		District Stayers			District Movers		
	Postal Code Switch	District Switch	Higher Income	Higher Ed	Distance	Higher Income	Higher Ed	Distance
1996	80.4	37.2	35.4	34.6	3.9	57.4	44.5	18.2
1997	78.1	37.0	35.3	36.4	3.8	54.2	47.1	18.2
1998	78.0	34.7	38.5	34.9	3.8	56.7	48.4	17.8
1999	77.0	36.0	37.7	35.2	3.7	55.3	47.9	18.8
2000	75.8	34.5	37.7	34.7	3.6	56.0	49.4	18.0
2001	76.4	33.8	38.5	36.6	3.7	58.4	47.7	18.0
2002	73.8	33.9	37.4	37.2	3.8	56.9	47.1	17.7
2003	72.5	34.3	36.6	35.0	3.7	54.5	46.8	18.2

All numbers in this table are percentages, except distance which is in kilometers. Postal code switch means the student switched both postal code and school. District switch is analogous. Columns 3-5 calculate statistics for all students that switch schools but not districts. The remaining columns are for students that switch schools and districts. Higher income means that the student moved to a school that is located in a higher income neighbourhood. Higher Ed means the student moved to an area with a higher proportion of people having some university education. Distance is the distance between schools.

Table 3.5: Characteristics of Schools by Observed Test Score Level

	Below Median Test Score	Above Median Test Score
% Male	51.7	51.2
% Aboriginal	5.7	2.6
% Non-English	34.9	26.3
% Special Ed	7.2	6.0
% French Immersion	4.8	6.2
% ESL	22.5	15.0
Household Income	59053.9	69464.4
S.E. of Household Income	1170.3	1795.8
% Education < Gr 9	8.1	5.6
% Education University	25.1	33.2
% 1-Parent Families	16.4	14.4
% 25+ Unemployed	6.6	5.4
% Immigrant	34.3	34.1
% Visible Minority	34.7	30.7
Avg Rooms per Dwelling	6.4	6.6
Avg Value per Dwelling	248875.9	311789.1
% Dwellings Owned	66.0	68.5
% Mobile Last Year	15.8	15.1
% Mobile Last 5 Years	47.1	45.0
Principal Tenure (years)	3.5	3.5
Student Teacher Ratio	15.2	15.8
% Numeracy Score = 0	8.2	6.9
% Reading Score = 0	8.0	6.9
% Excused from Numeracy	4.7	3.9
% Excused from Reading	4.8	4.1
% Public	92.6	82.8

Income and dwelling value are in 2000 dollars. Principal tenure is in years. The data used to produce this table are at the school level, and is pooled over the school years 1999 - 2003. Below median test score means the school scored in the lower half of all schools in the sample in that year. Census data is linked to school data at the Forward Sortation Area level, which is a larger area than the link used to match students to their home neighbourhoods.

Table 3.6: The Initial Effect of Increased Test Scores on the Probability of Switching Out of a School

	(1)	(2)	(3)	(4)
Score	-0.404***	-0.135***	-0.058	0.189
	(0.046)	(0.051)	(0.044)	(0.128)
Score × d^{2000}	-0.180***	-0.134***	-0.121**	-0.100*
	(0.051)	(0.051)	(0.050)	(0.051)
d^{2000}	-0.090	-0.147	-0.102	-0.000
	(0.183)	(0.182)	(0.180)	(0.173)
Constant	10.895***	8.156	7.238	17.335**
	(0.276)	(13.514)	(12.418)	(6.804)
R^2	0.008	0.012	0.028	0.035
N	141119	141119	141119	141119
Mean Dependent	7.898	7.898	7.898	7.898
Controls				
Time Invariant School	No	Yes	Yes	No
Time Varying School	No	Yes	Yes	Yes
Student	No	No	Yes	Yes
School FE	No	No	No	Yes

*p<0.10, **p<0.05, ***p<0.01. Dependent variable is an indicator for voluntarily switching schools between t and $t+1$. Data from years 1999 and 2000 and grades K-4 are included in these regressions. All coefficients are estimated by the Linear Probability Model. Only coefficients of interest are presented; full results available upon request. Score is the standardized student test score aggregated to the school level. $d^{2000} = 1\{year >= 2000\}$. Note that the dependent variable has been scaled so that the coefficients are percentages, not decimals.

Table 3.7: The Effect over Time of Increased Test Scores on the Probability of Switching Out of a School

	Raw Test Score		Smoothed Test Score	
	(1)	(2)	(3)	(4)
Score	0.160***	0.170***	0.105	0.109
	(0.051)	(0.050)	(0.068)	(0.066)
Score × d^{2000}	-0.120***	-0.085*	-0.145***	-0.098**
	(0.045)	(0.051)	(0.043)	(0.049)
Score × d^{2001}		-0.012		-0.011
		(0.064)		(0.066)
Score × d^{2002}		-0.063		-0.123*
		(0.071)		(0.068)
Score × d^{2003}		-0.018		0.028
		(0.076)		(0.066)
d^{2000}	0.719***	-0.042	0.724***	-0.043
	(0.152)	(0.175)	(0.151)	(0.176)
d^{2001}		1.041***		1.048***
		(0.197)		(0.197)
d^{2002}		0.018		0.032
		(0.245)		(0.246)
d^{2003}		0.203		0.191
		(0.292)		(0.297)
Constant	14.479***	16.193***	14.521***	16.325***
	(2.843)	(2.791)	(2.826)	(2.772)
R^2	0.034	0.034	0.034	0.034
N	357117	357117	357117	357117
Mean Switch	8.620	8.620	8.620	8.620
Controls				
Time Invariant School	No	No	No	No
Time Varying School	Yes	Yes	Yes	Yes
Student	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes

*p<0.10, **p<0.05, ***p<0.01. Dependent variable is an indicator for voluntarily switching schools between t and $t+1$. Data from years 1999 and 2003, and grades K-4 are included. All coefficients are estimated by the Linear Probability Model. Only coefficients of interest are presented. Score is a student-level z-score. $d^t = 1\{year >= t\}$, so coefficients on dummies are *increments* compared to $t-1$. Dependent variable is scaled so that coefficients are percentages. Raw test score is the test score for the school in that particular year. Smoothed test score is a two-year moving average between $t-1$ and t.

Table 3.8: Heterogeneity in the Effect of Increased Test Scores on the Probability of Switching Out of a School

	(1)	(2)	(3)	(4)
Score	0.205***	0.153***	0.158***	0.084
	(0.061)	(0.051)	(0.051)	(0.067)
Score × d^{2000}	-0.128***	-0.118***	-0.114**	-0.108*
	(0.048)	(0.045)	(0.044)	(0.064)
d^{2000}	0.659***	0.638***	0.473***	0.285*
	(0.159)	(0.160)	(0.181)	(0.166)
# Pub Schls × Score × d^{2000}	0.000			
	(0.005)			
# Priv Schls × Score × d^{2000}	0.009			
	(0.016)			
Avg Nearby Schools × Score × d^{2000}	0.023			
	(0.020)			
Income × Score × d^{2000}		0.001		
		(0.001)		
%Dropout × Score × d^{2000}		-0.051		
		(0.065)		
Score × Score × d^{2000}			0.019**	
			(0.007)	
Ability × Score × d^{2000}				0.004
				(0.005)
Non-English × Score × d^{2000}				0.018
				(0.094)
Constant	12.129***	11.909***	12.148***	11.126***
	(1.681)	(1.689)	(1.681)	(1.944)
R^2	0.034	0.034	0.034	0.034
N	357117	357117	357117	199597
Mean Switch	8.620	8.620	8.620	8.620
Controls				
Time Invariant School	No	No	No	No
Time Varying School	Yes	Yes	Yes	Yes
Student	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes

*p<0.10, **p<0.05, ***p<0.01. The main details are as defined in the previous Table. Average of nearby schools refers to the average score of all schools within 5 km of the school under study. Ability is the test score of the individual student, no matter when they took the test. The decrease in the sample size in column (4) is the result of having only a subsample of students whose individual test scores are observed in the relevant years.

Table 3.9: The Effect of Test Scores on Percentage Change in Kindergarten Enrollment

	Raw Test Score		Smoothed Test Score	
	(1)	(2)	(3)	(4)
Score	-0.003	-0.003	0.002	0.002
	(0.005)	(0.005)	(0.006)	(0.006)
Score × d^{2000}	0.005	0.010	0.007	0.008
	(0.005)	(0.008)	(0.005)	(0.008)
Score × d^{2001}		-0.016		-0.012
		(0.011)		(0.010)
Score × d^{2002}		0.017*		0.021**
		(0.010)		(0.010)
Score × d^{2003}		-0.005		-0.011
		(0.009)		(0.010)
d^{2000}	0.016	0.001	0.015	0.001
	(0.020)	(0.029)	(0.020)	(0.030)
d^{2001}		-0.023		-0.025
		(0.031)		(0.031)
d^{2002}		0.052*		0.050*
		(0.030)		(0.030)
d^{2003}		0.031		0.034
		(0.030)		(0.031)
Constant	0.064	0.080	0.061	0.073
	(0.094)	(0.095)	(0.095)	(0.095)
R^2	0.101	0.113	0.103	0.115
N	2070	2070	2070	2070
Mean Switch	0.073	0.073	0.073	0.073
Controls				
Time Invariant School	No	No	No	No
Time Varying School	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes

*p<0.10, **p<0.05, ***p<0.01. Dependent variable is $\frac{enr_t - enr_{t-1}}{enr_{t-1}}$. Regressions are OLS at the school level. $d^{2000} = 1\{year >= t\}$, so coefficients on the dummies are the *increment* in the intercept compared to the previous year. Raw test score is simply the test score for the school in that particular year. Smoothed test score is a two-year moving average between $t-1$ and t. The coefficients reported reflect the effect of an increase of 1/10 of a standard deviation, or about 2 percentage points.

Table 3.10: Conditional Logit Effects of Test Scores on Kindergarten School Choice

	All Schools	Schools <= 5km
Score	0.018**	0.021**
	(0.007)	(0.008)
Score × d^{2001}	0.026***	0.026***
	(0.008)	(0.009)
Income × Score × d^{2001}	0.000	0.000
	(0.000)	(0.000)
Income × Score	0.001***	0.001**
	(0.000)	(0.000)
%Dropout × Score × d^{2001}	-0.011*	-0.009
	(0.006)	(0.007)
%Dropout × Score	-0.015***	-0.018***
	(0.005)	(0.006)
Non-English × Score × d^{2001}	-0.034***	-0.031***
	(0.011)	(0.012)
Non-English × Score	-0.033***	-0.032***
	(0.009)	(0.010)
% Non-English	-0.012***	-0.011**
	(0.004)	(0.004)
% Aboriginal	-0.228***	-0.214***
	(0.013)	(0.014)
% Lone Mother Families	0.000	0.121**
	(0.043)	(0.051)
% Dropout	0.437***	0.334***
	(0.053)	(0.067)
% Unemployed 25+	-0.046***	-0.019*
	(0.008)	(0.010)
% Visible Minority	-0.033**	-0.016
	(0.013)	(0.017)
Household Income	0.021***	0.025***
	(0.002)	(0.002)
% Mobile Last Year	0.768***	0.599***
	(0.047)	(0.053)
Distance	-1.564***	-2.059***
	(0.022)	(0.022)
N	2261600	745104
Log Likelihood	-45454	-35851

*p<0.10, **p<0.05, ***p<0.01. Only data from Vancouver and Burnaby are used. Column 1 uses all schools as the individual's choice set, and column 2 restricts to all schools within 5 km. Years 2000 - 2004 are used. The dummy d^{2001} is used because the test scores here are lagged twice.

Figure 3.1: Spatial Distribution of Test Scores

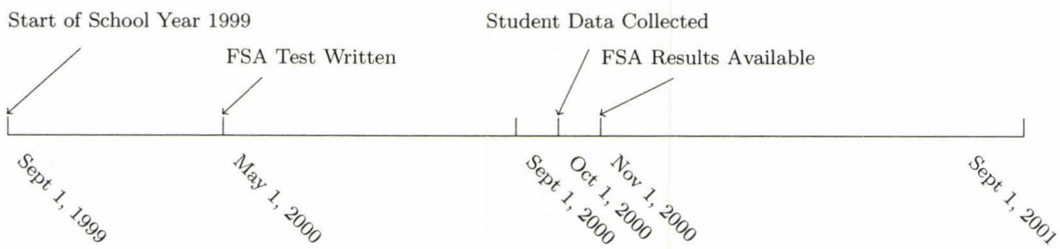


Figure 3.2: The Timing of Test Score Release and Data Collection

3.A Appendix: Linkage of Census Data

The BC administrative data do not contain any parental information, which is important information to have in an analysis such as this. The best available option is to proxy for these factors using data from the Canada Census at low levels of aggregation. Unlike the U.S. Census, the Canadian Census is taken once every 5 years. Statistics Canada distributes two forms to households: the short form asks basic demographic questions to 100% of the households in the country, while the long form asks a long list of detailed questions to 20% of the population. The responses are compiled and published at various levels of aggregations for public use. Currently, the lowest level of aggregation that Statistics Canada uses to publish the Census is called the Enumeration/Dissemination Area (EA/DA). EA is the older name for a small geographic area within which it would have been feasible for a Census counter to enumerate the people living in that area. The name, and boundaries, have since been changed to DAs, but both representations are similar geographic areas. They are designed to contain about 400-700 people.

Unfortunately, this data cannot be linked directly into the BC administrative data since it does not identify within which EA/DA a student lives. Instead this link must be done indirectly through the postal codes contained within the BC data. The Canadian postal code is a 6-digit string consisting of alternating numbers and letters. The first 3 digits are called the Forward Sortation Area, and indicate the province and sub-area where that the postal code represents. The full 6 digits represents a very small area in urban zones, sometimes as small as one side of the street on a city block. As such, it is a very precise indicator of residential location, and a perfect way to link students to EA/DAs.

Currently, Statistics Canada publishes the Postal Code Conversion File (PCCF), which links all Canadian postal codes to various levels of aggregation contained in the Census, including EA/DAs. Sometimes the geographic area of a postal code crosses the boundaries of more than one EA/DA. In these situations, Statistics Canada has included a "single link" indicator that defines the best match between postal code and EA/DA based on a method that identifies the amount of area a particular EA/DA covers within the postal code. For the purposes of matching, I use only those single links, and discard all other matches. Thus, the BC student data is matched to the PCCF, which is then matched to the Census data.

Because I have student data on a yearly basis and Census data on a semi-decennial basis, a natural question is which Census records to merge to each year? Currently, I only have access to the 1996 and 2001 Censuses, so the approach I take in this paper is to merge the 1996 Census data to all student records before 1998, and the 2001 Census to the remaining records. For this reason, there is no variation from year to year on the Census characteristics, and in general it is preferable to use fixed effects when possible.

3.B Appendix: Robustness Tables

Table 3.B1: The Effect of Increased Test Scores on the Probability of Switching Out using .75 Quantile

	(1)	(2)	(3)	(4)
Score	-0.376***	-0.096*	-0.019	-0.073
	(0.050)	(0.052)	(0.045)	(0.128)
Score × d^{2000}	-0.155***	-0.118**	-0.116**	-0.133**
	(0.054)	(0.053)	(0.052)	(0.052)
d^{2000}	0.985**	0.662	0.700	0.910**
	(0.454)	(0.450)	(0.443)	(0.434)
Constant	13.511***	10.106	7.533	18.312***
	(0.463)	(13.630)	(12.451)	(6.826)
R^2	0.007	0.012	0.028	0.035
N	141119	141119	141119	141119
Mean Dependent	7.898	7.898	7.898	7.898
Controls				
Time Invariant School	No	Yes	Yes	No
Time Varying School	No	Yes	Yes	Yes
Student	No	No	Yes	Yes
School FE	No	No	No	Yes

*p<0.10, **p<0.05, ***p<0.01. Dependent variable is an indicator for voluntarily switching schools between t and $t+1$. Data from years 1999 and 2000 and grades K-4 are included in these regressions. All coefficients are estimated by the Linear Probability Model. Only coefficients of interest are presented; full results available upon request. Score is the standardized student test score aggregated to the school level. $d^{2000} = 1\{year >= 2000\}$. Note that the dependent variable has been scaled so that the coefficients are percentages, not decimals.

Conclusion

In this thesis, I provide answers to some important, unanswered questions in the economics of education. I focus specifically on two topics that have generated a substantial amount of attention not just in the economics community, but in other academic circles and among the general public. The first two chapters address questions about the effects of school entry policy on student outcomes in early elementary school and in late high school. The last chapter switches gears and explores the extent to which providing parents with additional information about the quality of schools influences their school choice. Like much of the current research in this field, the majority of the empirical work contained within this thesis has been made possible by the availability of a high quality data set, which in this case comes from the province of British Columbia (BC).

In the first essay, I look specifically at the persistence in the test score differences among children in the same grade who vary in age. While the existing literature has established that older students perform much better than their younger classmates in early elementary school, it has not established to what extent this advantage lasts to the end of compulsory schooling. My main contribution is to provide estimates of this persistence from a consistent set of data, but I also provide a methodological contribution by comparing the status quo method to one that is arguably more robust to the influences of unobserved factors. I find that children who are one year older than their classmates perform substantially higher on numeracy, reading, and writing tests in grade 4, and continue to perform better but at a slightly lesser magnitude in grade 10. Furthermore, I find that the large differences observed in early elementary school persist at a stronger magnitude for girls than for boys. In terms of methodology, while comparing the status quo Instrumental Variables method to the Regression Discontinuity Design reveals similar estimates, it does raise the potential for a small amount of bias for the status quo.

The second essay looks deeper into *how* age affects student outcomes. Most research in the literature has interpreted age-related test score differences as the effect of entry age, but because students also vary in age at each stage of their schooling, the entry age effect cannot actually be disentangled from a test age effect. I contribute to the literature by separating these two effects using a group of students entering school around the time of BC's dual entry policy, and by focusing in particular on the differences in each effect across gender. Such a separation allows both parents and policy makers to make more informed decisions as they attempt to maximize student performance in school. I find that age-related differences in the probability of repeating grade 3 are due to test age, while age-related differences in grade 10 test scores are due to entry age. Both the entry age effect and the test age effect are at least twice the magnitude for boys.

In essay 3, I assess whether providing more information to parents allows them to make more informed decisions about which school their child attends. The related literature looks exclusively on the effect of repackaging existing information in a simple way on parent school choices, whereas my goal in this paper is to examine how parents react to an entirely new source of information. I contribute to the literature first by showing theoretically how parents react to such information, and second by providing empirical support for the model's conclusions. Using the theoretical model, I show that parents will use new information on school quality if they believe that information is not too noisy, and if they have poor prior information. Furthermore, as more observations on the new source of information are revealed, parents will make increasing use of it. When I test these results empirically using a set of test scores revealed suddenly to parents in BC, I find that parents are more likely to switch out of schools that perform poorly on the test. I also find that this negative relationship is stronger when more test scores are available, and that it varies by the education level of the neighbourhood and whether the student speaks English at home.

While I have only provided answers to a small number of the outstanding questions in the field, the results of this thesis provide a substantial amount of information that help to inform both parents and policy makers about the effects of what would otherwise seem like innocuous policy. In particular, the results from the first two essays that age-related differences in test scores can persist for so long, that these effects are rooted in differences in entry age, and that there is a large variance across genders, should help guide corrective policy. For those interested in increasing performance of public schools, the results of the last essay suggest that providing more information on school quality may have some effect on increasing the competition that schools face, but this effect is not likely to be large, and funds allocated to providing more information may have better uses elsewhere.