THE UNIVERSITY OF CHICAGO

IMPROVING THE DISTRIBUTION OF EDUCATIONAL OPPORTUNITIES

A DISSERTATION SUBMITTED TO
THE FACULTY AT THE IRVING B. HARRIS
GRADUATE SCHOOL OF PUBLIC POLICY STUDIES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
ALVARO SEBASTIÁN GALLEGOS VALLEJOS

CHICAGO, ILLINOIS
JUNE 2016

A Valeria, mi bálsamo y motor

…History's so strong (x4).

Red Hot Chili Peppers (2003)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am especially indebted to professor Robert J. LaLonde for believing in me from the very beginning. I would also like to acknowledge my former mentors in Chile, Pilar Romaguera and Dante Contreras, and my friends Andrés Hojman and Pablo Celhay for their decisive support at the start of my studies at The University of Chicago.

I enjoyed my doctoral studies to a full extent. It is difficult to single out specific moments, but I will never forget how hard we studied for our core courses with Pablo, Bocar, Rebbeca and Carolina; the first class with Gary Becker, game theory with Roger Myerson, the clarity of Steven Levitt, and the never-ending lectures by Jim Heckman –who used to order dinner to keep us in the classroom for five continuous hours and more.

The teaching assistant (TA) moments come to my mind too; I learned tons being TA for my advisors Dan Black and Robert LaLonde, and in particular I cherish TAing for Ofer Malamud. I learned much from him, both being his TA and from his permanent advise on my own research. I also thank Susan Mayer and Ariel Kalil, who have challenged me to engage in rigorous and policy relevant research. Working with them has been immensely enlightening and a critical component of my formation as a scholar. Finally, I especially thank Cynthia Cook-Conley; no student at Harris can survive the Ph.D. experience without her constant and unconditional support.

I also thank the warm support from my family and friends, who were always there for the difficult moments. Last but hardly least, I am indebted to my wife Valeria for her support, love and patience; this is also her accomplishment.

ABSTRACT

This dissertation is composed by four chapters studying the accumulation of human capital at different stages of the life cycle. Each chapter concludes with policy recommendations aiming to improve the distribution of educational opportunities, from preschool to adult life.

Parent engagement with their children plays a key role in their eventual economic success. Chapter 1 reports results from a randomized field experiment designed to increase the time that parents in subsidized preschool programs spend reading to their children using a reading application. The treatment, which included three behavioral tools (text reminders, goal-setting, and social rewards) increased usage of the reading application by one standard deviation after a six-week intervention. The findings suggest substantial promise for the application of behavioral tools to parenting interventions that promote investments in children's human capital.

Chapter 2 examines whether the initial benefits of delayed school enrollment persist in the long run. Regression-discontinuity estimates, based on birth dates cutoffs, show that there are initial positive effects of a one-year delay on grades, test scores, attendance and persistence, but they decrease monotonically as students grow older. Those effects vanish by the end of high school and have no effect on the likelihood enrolling in any higher education institution. These findings indicate that delaying school enrollment in childhood does not have significant lasting effects and thus provide evidence against the counterintuitive recommendation from previous literature that children should start school later.

Chapter 3 examines whether test-score-based university admissions discard high-ability, low-income applicants. I use data from a policy that offered slots to top high-school students from disadvantaged backgrounds with scores below admission cutoffs. I find that beneficiaries (i) enrolled in university-major combinations with higher expected wages compared to their next-best alternatives; (ii) would not have fared significantly better in less selective majors; and (iii) are equipped with skills not measured by test scores that enable them to outperform their peers in their same majors. The findings suggest that admissions may be improved on

efficiency and equity simultaneously.

Chapter 4 links data on educational attainment for three generations in Chile. The main findings indicate that grandparental education influences grandchildrens schooling even after taking the parental factor into account. Accordingly, standard two-generation estimations over-predict intergenerational mobility over three generations. We find that (i) upward schooling mobility has moderately increased with younger cohorts, (ii) parents place of origin matters for upward mobility, as shown by heterogeneity in mobility across geographical regions, and (iii) having more educated same-sex ancestors matters more for women and suggests that gender-related social roles may be passed along generations within families. All in all, the results suggest that family background effects can be longer lasting than previously believed, affecting the endowments and idiosyncratic capabilities of children.

CHAPTER 1

USING BEHAVIORAL INSIGHTS TO INCREASE PARENTAL ENGAGEMENT

(With S. Mayer, A. Kalil and P. Oreopoulos)

1.1 Introduction

A large research literature documents the substantial differences in parental engagement between advantaged and disadvantaged families and the importance of parental engagement to children's future economic outcomes. Yet, existing programs intended to increase parental engagement have at best met with little success and sometimes at a high per child cost.

To address these issues we developed an experimental intervention designed to increase the amount of time that low-income parents spend reading to their children using an app on an electronic tablet. The intervention is based on principles from behavioral science that have been shown to be successful in changing behaviors such as financial savings (Meier and Sprenger 2010), smoking cessation (Rodgers et al 2005), children's literacy scores (York and Loeb 2014), college attendance (Castleman and Page 2014), and weight loss (Chabris et al. 2008). Our objective measure of how much time parents spend reading to their children combined with survey data from the parents allows us to assess the potential importance of several possible explanations for fact that disadvantaged parents seldom read to their children.

Over the past few decades, a growing research literature has demonstrated that decision-making is influenced by a variety of cognitive biases. A cognitive bias is a type of thinking that occurs when people focus too much on the present or adhere too much to routine such that opportunities for greater expected lifetime welfare are missed (Gilovich, Griffin, and Kahneman 2002; Lavecchia, Liu, and Oreopoulos 2014). For example, a substantial research literature has established that, in a variety of settings, individual responses deviate from those predicted by a time-consistent intertemporal utility model that assumes a constant discount rate (Stanovich, West, and Toplak 2012): future gains are discounted more than future losses; small changes to outcomes are discounted more than large changes (Frederick, Lowenstein, and O'Donoghue 2002); small probability events, when emphasized, are discounted less than when not emphasized; and responses depend on context, emotional state (Rick and Loewenstein 2008) and perceived social identity (Benjamin, Choi, and Strickland 2010; Galvan 2012). These cognitive biases result in suboptimal decision-making. In the case of parenting, decisions made today have associated payoffs in the distant future, expressed in terms of the child's future economic and social success. In this behaviorally-informed intervention we were especially focused on the cognitive biases in decision-making resulting from discounting the future.

Our main results indicate that the treatment increased the number of minutes that parents read to their children using an electronic application by one standard deviation, which equals twice as much reading time compared to the control group mean. The increase was much greater for parents with a high discount rate for the future as measured on a time preference task that we administered. Our findings are suggestive of substantial promise for the application of behavioral tools to parenting interventions that promote investments in children's human capital.

The remainder of this paper proceeds as follows. Section 1.2 provides background on differences between advantaged and disadvantaged parents in both children's skills and parental engagement. Section 1.3 provides details about the intervention. Sections 1.4 and 1.5 describe the results and present robustness checks, respectively. Section 1.6 concludes.

1.2 Background

1.2.1 Socio-economic Gaps in Children's Skills and in Parenting

Substantial differences between the skill development of economically advantaged and disadvantaged children emerge well before the start of formal schooling. For example, pre-school age children from families in the poorest income quintile score on average at the 34th percentile in a test of literacy compared to children in the richest quintile who score at the 69th percentile (Waldfogel and Washbrook 2011). These gaps remain fairly constant through the school years (Heckman 2008, Cunha et al 2006) suggesting that most of the effect of family background occurs before age six. In addition, conventional measures of school quality (teacher/pupil ratios and teacher salaries) have small effects on creating or eliminating gaps after the first few years of schooling (Carneiro and Heckman 2003; Cunha and Heckman 2007).

Children spend the vast majority of their time from birth to adulthood in a family or other environment selected by parents and not in schools. For example children in the United States will spend only about 15-18 percent of their waking hours in school between birth and age 18.[1] Finally the intergenerational persistence of schooling reflects parent-child similarities in traits influencing educational attainment, such as orientation towards the future, sense of personal efficacy, work ethic, and other characteristics sometimes referred to as "non-cognitive skills" (see e.g. Cunha and Heckman 2007; Carneiro and Heckman 2003; Heckman and Kautz 2012). These skills are largely shaped by family influences and not by schools.

Children's experiences prior to the start of primary school clearly play a large role in their

1. In the United States the average length of the school day was 6.4 hours in the 2009-2010 school year and the average length of the school year was 180 days according the National Center Educational Statistics, Schools and Staffing Survey for that year. If we assume that this number is constant over all the years that a child attends school and that he number of years including kindergarten is 13, then (6.4*180)13 =14,976 is the total number of hours spend in school over the first 18 years of a child's life. If we assume that children average 16 hours awake each day the number of waking hours in the first 18 years is 18(16*364) = 104,832. Thus 14,976/104,832 = 13.2 percent of a child's waking hours in the first 18 years is spent in school.

futures. This insight has led to an emphasis on pre-kindergarten programs for disadvantaged children. Benefit-cost assessments of the Perry Preschool and the Abecedarian program and recent research in brain science showing the importance of early brain development support the notion that early child development is crucial to children's eventual success (Knudsen, Heckman, Cameron and Shonkoff 2006). However, although small-scale model preschool programs have resulted in important gains in children's educational attainment, earnings, and lower rates of getting in trouble with the law, Head Start, which is the primary preschool program for disadvantaged children in the United States, has produced at best small benefits for children. This is partly because even children enrolled in full-day preschool programs will spend only about 22% of waking hours from the age of 3 to 5 in preschool, and virtually no time at all between the ages of 0 and 3. Adding a program like Early Head Start will still leave parents with much more influence on children's development than school settings. Consequently attempts to substitute school-based programs for parents' engagement with their children cannot alone succeed in diminishing the effects of family background unless they are very intensive, extensive and expensive.

Across disciplines, dozens of studies have demonstrated many differences in the way advantaged and disadvantaged parents engage their children and how these differences matter to children's adult success. Among other things, advantaged parents spend more time, in particular on educational activities, with their children (Guryan, Hurst, and Kearney 2008; Kalil, Ryan, and Corey 2012) and produce more cognitively stimulating home learning environments (Harris, Terrel, and Allen 1999). In addition, the gap in the amount of time that parents spend with their children overall and in educationally relevant activities has widened over the last 20 years because the rate at which highly educated mothers have increased the amount of time they spend in educational activities with their children has exceeded the rate at which mothers with less education increased such time with their children (Altintas 2012; Ramey and Ramey 2010; Hurst 2010).

In observational data these types of parenting practices explain large shares of the gaps between poor children and their more advantaged counterparts in cognitive and non-cognitive skills (Waldfogel and Washbrook 2011). In several papers James Heckman and his colleagues conclude that more engaged parents have greater success in producing both cognitive and non-cognitive skills in their children, and that both types of skills are crucial to social and economic success (see e.g. Cunha, Heckman, Lochner, and Masterov 2006; Heckman and Masterov 2004).

Other research suggests that spending time with a child has a direct and causal effect on children's cognitive test scores (Villena-Rodan and Rios-Aguilar 2012) and that the amount of time that parents spend in educational activities with their children is the most productive input for cognitive skill development (Fiorini and Keane 2014).

The intervention reported in this paper focuses on one particular type of parental engagement, namely reading to one's children. It is widely believed that reading to young children helps develop their literacy skills including word knowledge and recognition. Observational research supports this belief and suggests that parental reading to young children is also associated with greater numeracy and social skills (see e.g., Kloosterman, Notten, Tolsma and Kraaykamp 2011, Mol and Bus 2011, Hale et al. 2011, and Raikes et al. 2006. See also Kolb and Ours 2014 for a review of the observational literature on the relationship between parental reading and children's cognitive skills.). The few attempts to estimate a causal relationship between parental reading and children's cognitive skills also conclude that parental reading time increases cognitive test scores. For example, Price (2010) uses NLSY data to estimate a model that uses birth order to instrument parental time reading to children. According to Price (2010, 2012), second-born children get only slightly less total time with their parents compared to first-born children, but that time spent reading to or with the child is crowded out by other activities such as watching television. He also argues that birth order does not affect cognitive skill directly but only through differences in parental engagement. Using the birth order instrument and looking at families with two children, Price concludes that an additional year of daily motherchild reading would increase children's reading test scores in the early school grades by on average 41 percent of a standard deviation on the PPVT. Kolb and Ours (2014) using a similar instrument find that additional parental time reading to four and five year old children has a substantial and statistically significant effect on both literacy and numeracy skills and that the effects remain at least until children are ten to eleven years old. They experiment with a variety of instruments and with estimates using propensity score matching. Their results are robust to these different approaches and to numerous sensitivity tests. They find that on a standardized test of cognitive skills boys who are read to 3-5 days a week score just over a half a standard deviation higher than boys who were read to fewer than three days per week. Boys who were read to 6-7 days per week score just under one full standard deviation higher than boys who were read to less than three times per week. The results for girls are qualitatively similar.

Although both general social beliefs and the research evidence support the idea that reading to one's children is very important for their cognitive development, a substantial numbers of parents do not read to their children on a regular basis. In 2007, 16% of all parents of a child aged 3-5 years and 29% of parents living below the poverty line reported that neither they nor other family members read to their children at least three times per week (Noel, Stark, and Redford 2015).

Many parenting interventions have tried to increase the amount of time that disadvantaged parents spend in educationally enriching activities and especially reading with their children. However, most of these programs have had modest success at best. For example, many practitioners and policy makers believe that home visiting programs are the best way to improve parental engagement in disadvantaged families. Yet evaluations of these programs find little improvement in parent engagement or other aspects of the home environment (Olds, Henderson, and Kitzman 1994; Wagner, Spiker, and Linn 2002). This is important because the average cost to serve a family for 45 weeks in a home visiting program is about $6,500; the Nurse Family Partnership program is on average more expensive and can cost up to almost $14,000 for each participant (Zaveri, Burwick, and Maher 2014). By comparison, the Perry Preschool Program, which is widely upheld as a model preschool program, had effect sizes on arithmetic achievement at age 14 equal to 34% of a standard deviation but at a cost of $20,500 (in 2013 dollars) for each participant (Schweinhart et al. 2005).

The lack of parental participation and persistence in home visiting programs is an important problem that is also found with other programs intended to increase parental engagement with their children. For example, almost half of families left the home-based Early Head Start program before their child was 30 months old (the program was designed to last until the child was 36 months old), and more than a third dropped out before they had been enrolled for 18 of the 36 months (Roggman, Boyce, and Innocenti 2008). Only 40% of families stayed enrolled in Early Head Start home-based programs until graduation or transition. Other programs intended to change parenting behaviors experience similarly high attrition rates (Baker, Piotrkowski, and Brooks-Gunn 1999; Gomby 1999). Even if parents are motivated to invest in their children, many may be less motivated to participate in programs intended to increase their investment.

1.2.2 Obstacles to Parent Engagement: Insights from Behavioral Science

Multiple theories try to explain the difference in parental engagement with their children. Yet we have remarkable little empirical evidence on the salience of these theories. An important contribution of this paper is to try to understand the mechanisms influencing parental engagement in one particular important behavior, namely reading to their young children.

Perhaps the dominant explanation for differences in parental engagement by family advantage is that disadvantaged parents expect a lower return for the time they spend with their children. Agee and Crocker (1996) use parent's decision about whether to treat high levels of lead found in their children's blood as an instrument for the parents' discount rate for child investments. They find that less educated and lower income parents discount their investments at about twice the rate as more advantaged parents and that both have discount rates that are high compared to other investments. Cunha, Culhane, and Elo (2013) find that the median subjective expectation about the elasticity of child development with respect to investments is between 4% and 19% for mothers of very young children who receive Medicaid. In comparison, the "true" elasticity estimated from CNLSY/79 data is between 18% and 26%. This means that on average disadvantaged parents underestimate the return to their investment in their children but still estimate a substantial positive return. Differences in expected returns could account for some of the difference in the amount of time that parents spend with their children, although a lower expected return could lead to either less or more investment.

It is also possible that the difference between advantaged and disadvantaged parents in time investments in their children stems from disadvantaged parents' experiencing less positive affect from interacting with their children. But Kalil, Mayer, Delgado, and Gennetian (2015) show that in fact mothers with less than a college education report more enjoyment from spending time with their children than do mothers with at least a college education. Disadvantaged parents could also be time constrained. However, as we noted above factors such as work hours or number of children account for little of the SES-based differences in time investments in children.

It could be that disadvantaged parents engage their children less because they lack information about how to invest in educational activities, or that this information is just not salient enough for parents to remember to make use of it throughout the hectic day. York and Loeb (2014) provide some evidence for this using a text-messaging program for parents of preschoolers to provide quick tips of specific actions parents might take take to increase their child's literacy skills. Their 8-month intervention of 3 texts per week resulted in an increase in students' scores on an assessment of early literacy of 0.18 standard deviations at the end of the intervention (with a standard error of 0.104). Parental behavior corresponding closely with the content of the text messages also increased, though the increase was also estimated imprecisely and self-reported, and the study found no change in reading to the child.[2]

Another explanation for the gap in parental engagement between advantaged and disadvantaged families is that disadvantaged parents have less time and resources to engage their children. Evidence from time use data shows that even when researchers control the number of and ages of children in the home, whether the parent is married, and how much time the parent works, the differences by education in the amount of time that parents spend with their children remains. Furthermore, as we describe below few parents in our sample say that they lack the time to engage with their children and almost none say that they do not read because they have nothing to read.

Our approach is based on a rapidly growing research literature in psychology and behavioral economics that has focused on understanding why adults often fail to do things that they know they should do such as exercise, save, eat healthy foods, and how to change these behaviors. A leading explanation focuses on problems related to discounting the future such as impatience and present-bias (Thaler 2015). It is well known that future outcomes are under-emphasized (discounted) relative to immediate outcomes. This means that it is hard for people to give up things they enjoy today for the (under-valued) future and the result is problems of self-control. In the context of parenting it is hard to give up leisure today for a distant return on children's human capital. Discounting the future induces procrastination and impatience, which affects activities in many domains, such as in financial choices (savings and credit card borrowing, e.g. Meier and Sprenger 2010; Eckel, Johnson, and Montmarquette 2005), health-related behaviors like dieting and exercising, smoking and nutrition intake (e.g., Chabris et al. 2008), or education-related behavior like investments in human capital, studying, or behaving at school (Sutter et al. 2013; Castillo et al. 2011).

Some research has documented that low-income individuals discount the future more

2. The differences in home literacy activities between the treatment and control group are usually quite small in absolute terms. For example although the difference between the treatment and control group in the number of times that parents report telling stories to their children is .22 standard deviations and statistically significant at p=.05, in absolute magnitude the control group told stories 2.85 times per week and the treatment group told stories 3.03 times per week.

heavily than higher income individuals (Golsteyn, Grnqvist, and Lindahl 2013; Lawrance 1991; Dohmen et al. 2010). Other research suggests that this might help account for differences in child rearing behavior. For example Pabilonia and Song (2013) find that even after controlling for parental differences in income, employment, and education, single mothers who are more present-oriented spend significantly less time with their children overall and less time engaged with their young children in educational activities. Their children also have lower scores on reading comprehension tests.

Differences in time preferences may arise because disadvantaged parents face a host of stressors such as income instability, child care problems, car breakdowns, and the like, all of which place cognitive and emotional demands on their attention in the present and leave little energy for thinking about the future (Spears 2011; Mani et al. 2013; Gennetian and Shafir 2015). The cognitive load associated with such stressors may result in inattention and distraction that result in less emphasis on the future and impatience today. Differences in the focus on the present may arise for other reasons and may contribute to disadvantaged parents having lower education and income. Impatient individuals may prefer jobs with flat wage paths, as opposed to careers that promise high wages only after a period of training or education. Research has shown that certain behavioral tools such as commitment devices, reminders, and social rewards can help overcome impatience and procrastination and lead to better decisions.

Behavioral scientists have proposed a number of approaches for addressing present-bias. A commitment is a pledge to carry out a specific behavior or take actions necessary to achieve a specific goal (see Ashraf 2013; Bryan, Karlan, and Nelson 2010 for recent relevant studies). People are strongly motivated to be consistent with their own past actions, especially actions taken publicly. For this reason, when a commitment formalizes a person's pledge to do something or achieve an objective, it increases the chance that he or she will do it by increasing the psychological cost of not doing the behavior to which the individual committed. Research suggests that even the act of writing down a commitment to doing an activity can increase the likelihood of actually doing it (Giné, Karlan and Zinman 2010). Commitment devices reduce procrastination and impatience by focusing attention on the actions that the individual pledged to take.

Reminders can also reduce procrastination and impatience by inducing people to attend to tasks by refocusing attention on the task. A reminder can change time allocation today by focusing attention on the relationship between future outcomes and current choices. Text messages are the most common way to communicate reminders and they are now a common feature of many interventions (Richburg-Hayes et al. 2014). Castleman and Page (2013) and Bergman (2015) also show that text message reminders to parents and adolescents can improve educational outcomes.

Social incentives are designed to shift preferences by increasing the utility of specific behavior, often by increasing their monetary value but sometimes by increasing the cost of alternative behaviors. While a large research literature documents the importance of monetary incentives for changing behavior, a growing body of research also documents the importance of nonmonetary or social incentives for changing behavior. Nonmonetary incentives may be at least as effective as monetary incentives at changing a behavior, such as parents engaging with their children, that is considered pro-social or normatively desirable (Gneezy and Rustichini 2000; Benabou and Tirole 2003, 2006; Besley and Ghatak 2005; Dixit 2002). Social incentives work by both increasing the utility of the rewarded behavior and, like commitment devices, reducing procrastination and impatience by focusing attention on the rewarded behavior.

1.3 The Parents And Children Together Intervention

We called our intervention the Parents And Children Together (PACT) program. Parents with children aged 3-5 enrolled in a subsidized preschool program (e.g. Head Start) in Chicago were given an electronic tablet to borrow for the six-week intervention. The tablet had an application pre-loaded on it that included over 500 children's books in English and

Spanish. Within each preschool center, parents were randomized to either a treatment group or control group. [3] [4] The treatment included three behavioral tools (a commitment device, reminders and a social incentive) plus information on the importance of parents' reading to their children. The intent of the intervention was to increase the time that parents spent reading to their children using the app on the electronic tablet.

We use what Bryan, Karlan, and Nelson (2010) call a soft commitment device. At the beginning of each week a research assistant asked each parent in the treatment group to set a goal for how much time he or she would spend using the reading app during the next week. The research assistant recorded this number in the "virtual goal keeper" (VGK), which is a web site that we developed for the project. At the end of the week the research assistant showed the parent a bar chart on the VGK (or sent a text message with the information) that displayed the parent's goal relative to how much the parent actually read. Parents were then asked to set a goal for the next week. This was repeated for every week during the intervention. During this time the research assistants did not interact with the parents except to show them their progress and get their new goal. While control group parents did not set a goal, they also briefly met with research assistants each week so that we could collect data from their tablets.

Parents in the PACT treatment group received a text message reminder every weekday. Text messages combined reminders to work toward the parent's goal of using the tablet to read to their child and messaging about the importance of reading. There were a variety of texts that repeated after the third week. Some examples were, "Are you working towards your goal this week? Remember that reading with your child is very important to your

3. We recruited parents when they dropped off or picked up their children at the centers. We spent between 6-10 hours in each center inviting parents to participate. Parents received different tablets depending on whether they were in the treatment or control group. The tablet number indicated the group to which they belonged. The recruitment process is explained in detail in the appendix.

4. Together the preschool programs served about 965 children at the time we began recruiting. Based on conversations with the staff of the preschools we estimate that about 15 percent of the 965 children were not eligible due to the language condition. We estimate that an additional 15 percent are siblings of another child in the center and only one sibling could be a focal child in PACT. Therefore, across the 8 centers, about 580 children were eligible to participate in PACT.

child's future." "What's your child's favorite book? Find some books your child loves and have fun reading them tonight!"

As a social reward if a parent met her reading goal she received a congratulatory text message, and if she viewed her reading minutes and goal on the Virtual Goal Keeper a cartoon bear did a celebratory dance. In addition, each week a text message went to all parents in a center saying, "This week in your group, the parent with [tablet number] did the most reading with their child! Congratulations to that parent!" The parent with that tablet number would therefore know that she had done the most reading but other parents would not know who that parent was.

We provided information on the importance of reading to treated parents while parents in the control group received 'placebo' information. In particular, tablets for parents in the treatment group included 13 videos and 2 .pdf documents about parenting, language development and the importance of parents' reading to their children. Tablets for parents in the control group included 14 videos and 1 pdf document about nutrition, health and dental hygiene. If a document was opened this information was stored on the tablet so we could identify whether parents opened the documents.

Parents in both the treatment and control group were asked to designate a "focal" child in case they had more than one child in preschool and they were instructed to use the reading app with this child. Parents were told that other children could be present when they used the app but that we were most interested in their use of the app with the focal child. We observed parents using the app to read to their child over a six-week period. When a book from the app was opened on the tablet an audio and video recording of the parent reading to the child automatically began and continued until the parent finished the book or closed the app.

We identified the number of minutes that parents read to their children by using the time stamps provided in the reading application. We implemented the intervention in two six-week rounds to minimize the number of electronic tablets that we needed to purchase.

The first round was in the autumn semester of the preschool programs and the second was in the winter semester. As we describe below, we implemented two short follow-ups with participants to test whether the effect of PACT persisted after the intervention ended. Parents whose primary language was either English or Spanish, who had a child enrolled in one of eight subsidized preschool program in Chicago, Illinois, who were willing to sign a consent form and a pledge to borrow, safeguard, and return an electronic tablet at the end of the program were eligible for the intervention.[5] The preschool programs were located throughout the city and have a racially and ethnically diverse population. [6] We collected data in 3 surveys. Survey 1 was administered when parents first received the electronic tablets. It asked parents to report general characteristics of the child, the household, and themselves. Survey 2 was administered in the middle of the intervention period and consisted of a time preference assessment that we discuss below. Survey 3 was administered at the end of the six-week intervention. It asked questions about parenting beliefs and practices.

1.4 Results

1.4.1 Do Parents Read More to Their Children?

A total of 169 parents participated in the PACT intervention. Among this group about half were randomized to the treatment (N=84) and half to the control (N=85) group. Of these, 160 took the first survey (80 treated, 80 control). Overall the parents participating in the PACT intervention are similar to a national sample of parents of children in Head Start programs on characteristics collected in the Head Start Family and Child Experiences Survey (FACES) survey except that a higher proportion of PACT families are Spanish speaking. Table 1.1 describes the main characteristics of the total sample and the treatment and control

5. All tablets had 6 instructional videos (each in Spanish and English) pre-loaded on them. Two of these were about how to use the reading app and the other four were an introduction to the PACT study, the consent form, and the tablet agreement form that participants were asked to sign.

6. A map of the City of Chicago with the geo-referenced preschool can be found in the appendix.

samples. No differences between the treatment and control group is statistically significant at the 5% level. Three differences were significant at the p=0.10 level, which is about what we would expect by chance.[7] The focal child from the treatment group was more likely to have been born early (p=.07). Treatment group parents were more likely to have a GED (p=.10) while control group parents were more likely to have some college (p=.10). Conditioning on these observed variables makes virtually no difference to the results.

Our main outcome of interest is the time parents spend reading to their children using the reading app on the tablet. [8] We also have information on related outcomes like the number of stories read by the parents; the average time spent reading each story, and whether parents read the same or different stories. We estimate the following equation:

$$Y = \alpha + \delta T + \varepsilon \tag{1.1}$$

where Y is an outcome variable, and T is an indicator for random assignment to our treatment. Our parameter of interest if δ, which equals the average difference for parents randomized to the treated group compared to those randomized to the control group. Below we add relevant covariates to this model to test the robustness of the results to their inclusion.

Column 1 in Table 1.2 shows results for the estimation of equation (1) with the number of minutes parents read to their children using the app as the outcome. On average, the treatment group read 88.3 minutes more than the control group over the six weeks of the intervention, more than doubling the amount of time spent read to the child using the app (an increase from an average of 63.3 to 151.7 minutes). This represents a one standard deviation difference. Column 2 shows that the average number of books read during the 6-week period

7. We compare differences in 22 variables in Table 1.1, so we expected 2.2 of them to be significant at the 10% due to chance.

8. Some parents may "outsource" the time investment to another family member, such as an older sibling, father, a grandparent or someone else. While in principle we could distinguish who was reading to a child in practice this became burdensome. In addition, there is no reason to expect that there would be differences between the control and treatment group in the tendency to outsource reading to others. Therefore, we counted all minutes in which the app was being used for actual reading. We spot checked videos to assure that there was both an adult and a child in them.

was 14.8 for the control group and 31.4 for the treatment group. In other words, parents in the treatment group read an average of almost one book per weekday compared to two or three books per week for the control group.[9] This is a large effect compared to York and Loeb's (2014) intervention in which at the end of the intervention treatment parents reported reading to their children 3.03 time while control parents reported reading 2.85 time. The last column in Table 1.2 shows that the treatment also increased the fraction of parents ever using the app to read a story to a child from 84 percent for the control group to 96 percent for the treatment group.

We conducted short-term follow-ups of parents who participated in PACT. One group of parents kept the tablets for an additional three weeks immediately after the intervention ended and another group got the tablets back for three weeks three months after the intervention ended. No behavioral tools were used during either of the three week follow-ups. We then compared the amount of time that treatment and control group parents spent using the reading app during the follow-up period. Although the effect sizes remained very large for the follow-ups, these results were inconclusive because of the small sample sizes and selection into the follow-up groups. These results are available in the appendix.

1.4.2 Do Parents Who Discount the Future Respond More to Reminders?

As noted, reminders and commitment devices are designed to overcome the procrastination that arises from discounting the future. If discounting the future reduces parents' engagement with their children and if these behavioral tools mitigate the discount rate on the future then we should observe that the parents who discount the future more read less to their children and respond more to the intervention. To explore this hypothesis, we implemented a time-preference task designed to measure parents' time discount rate.[10] We use the Convex Time

9. We also found that there was no statistically significant difference in the average time spent reading each story, or in the number of different books read by parents. These results suggest that the treatment effect occurred mostly through reading more books.

10. The time preference task was administered on electronic tablets after random assignment to parents in both the treatment and control groups. See the appendix for more on details, procedures and the ques-

Budget (CTB)[11] method (first introduced by Andreoni and Sprenger 2012) to estimate both a discount rate for parents and their present bias, which have been shown to best predict outcomes in previous research (Burks et al. 2012). We then examine whether parents with a higher discount rate read less and responded more to our treatment.

Parents were asked to choose between an amount of money that they could receive immediately or a different amount that they could receive later. In total, each parent answered 15 of these questions with four choices each (two inner solutions and two corner solutions), as in Andreoni, Kuhn, and Sprenger (2013). The first five questions gave parents the choice of receiving payments immediately or three weeks later, the next five questions gave parents the choice of receiving payments immediately or six weeks later, and the last five questions offered the choice of receiving payments in three or six weeks. Within each time horizon, each question presented an increasing price for the earlier payment.[12] The set of choices is shown in the appendix. The 3- and 6-week time horizons were chosen because the intervention lasted six weeks so the payout could be made during or at the end of the intervention. Comparing decisions with a delay of 21 days to those with a delay of 42 days allows us to identify the discount rate, or degree of "patience." Less patient (more patient) parents are defined as those with an estimated discount rate above (below) the sample median. The degree of patience among parents is balanced by treatment status.

Of the 169 participants who were eligible to participate in the time preference task, 112 participated. Of these, 59 were in the treatment group and 53 were in the control group and participation was not correlated with treatment status.[13] To understand how representative

tionnaire.

11. The CTB method identifies the curvature in the utility function over time by estimating a respondent's sensitivity to changing interest rates. The key in using this approach is to vary the implicit interest rate in the options presented across subsequent sets of options. The sensitivity to changing interest rates across the question sequences identifies the utility curvature; the time preference is identified through the stated preference over the timing of payments.

12. Decreasing the earlier payment is equivalent to increasing its price relative to the future payment.

13. 70% of the treated parents participated in the time preferences task, while 62% of the control group parents did. The mean difference of eight percentage points (pp.) has a standard error of 7.3 pp., and is therefore not significant at the .05 level.

of our original 169 parents these 112 parents are, we first repeat the treatment regression shown in Table 1.2 for the 112 parents who participated in the time preference task. The first column of Table 1.3 repeats the result from Table 1.2 and the second column shows the results for the 112 parents who participated in the time preference task. The coefficient on the treatment condition is 82.1 in that model compared to 88.3 for our baseline estimation (the difference between these coefficients is not statistically significant).[14] We also compared parents who did and did not participate in the time preference task on the characteristics shown in Table 1.1. On these 22 characteristics no difference between the two groups was statistically significant at p=0.05.

Column 3 in Table 1.3 shows that the treatment effect for less patient parents was 124.52 minutes and column 4 shows that the treatment effect for more patient parents was only 42.26 minutes. The constant indicates the amount of reading net of the treatment. It shows that more patient parents read 97.26 minutes compared to less patient parents who read only 68.33 minutes. Thus, absent the treatment more patient parents read more than less patient parents but the treatment has a greater effect for less patient than for more patient parents. These results are consistent with the theory that parents who discount the future more invest less in their children's future but that the behavioral tools in PACT help these parents more. Factors correlated with a parent's discount rate may account for the differences in the effect of PACT. For example, parents who experience "cognitive scarcity" (Shah, Mullainathan, and Shafir 2014; Spears 2011, Mani et al. 2013, Gennetian and Shafir, 2015) may both find it hard to attend to their child's needs to value the future. It is difficult to empirically separate these mechanisms.

14. The z-test for the difference is given by $z = (B1 - B2)/\sqrt{(seB1^2 + seB2^2)} = .203$, and hence non-significant.

1.4.3 Did Parents Learn from Goal-Setting?

Each week parents were asked, "How many days and how many minutes per day do you commit for your goal this week?" Anecdotally we know that parents in PACT attended to the goals that they set because they often told members of the research team that they were sorry that they did not meet their goal and expressed determination to meet the goal. In theory, goal setting both focuses attention on the behavior that is the object of the goal thus reducing procrastination or impatience and by imposing a psychological cost to not meeting the goal.

If not meeting the goal is psychologically costly then we would expect that over the six weeks of the intervention the difference between the goal and the actual amount of reading time would decline. This in fact happened. In week one the average actual reading time for parents in the treatment group was 23 percent of the average goal. By week 6 the average actual reading time was 65 percent of the average goal. The convergence mainly occurred because the goals decreased each week (from an average of 93 minutes in week 1 to an average of 49 minutes in week 6). The actual reading time varied from 26 minutes per week to 34 minutes per week but there was no clear trend in the amount of reading over the intervention period.[15] Although parents became more realistic about how much they would actually read in the coming week, after six weeks the difference between the goal and actual reading time was still sizable.

1.4.4 Did Information about Reading Matter?

We explore whether information about the importance of parental engagement and reading in particular was an important part of the treatment effect by examining the responses to the survey that was administered at the end of the intervention. As noted the electronic tablets for parents in the treatment group had videos and .pdfs on the importance of parental

15. The appendix describes the process for estimating the average weekly goals and the average weekly actual amount of reading time.

engagement and especially reading while the tablets for the control group had placebo videos and .pdfs. However, only 15 parents in the treatment group and 15 in the control group ever opened a video or .pdf during the six-week intervention. Informal discussions with parents suggest that they mainly opened these as part of their exploration of what was on the tablet and not to engage in the content.

In addition, the survey that we administered at the end of the intervention asked three questions about parental engagement that were the subject of the informational material on the treatment group tablets. These questions were on parents' expected return to the time they spend reading to their children, their beliefs about whose job it is to teach reading to their children, and whether they think their child's ability is set at birth or not.

If the information on the tablets made a difference we would expect parents in the treatment group to report a higher expected return to their time compared to the control group parents. In the survey at the end of the intervention we asked the following question: "If you were to spend 15, 30, 45 or 60 more minutes per day in educational activities with your child, how do you think your child's reading skills would compare to all other children of the same age across the United States, on a 0-10 scale?" Figure 1.1 shows that parents in both the treatment and control group believe that investing more time in educational activities would increase their children's skills. However, we found no statistically significant differences between treated and control parents on this measure.

In the same survey we also asked parents, "Whose job is it to teach math and reading to preschool children."[16] If parents think it is mainly the job of the preschool to teach math and reading it implies that they do not believe that they need to spend educational time with their child in order to improve these skills. But only 1 percent of parents in the treatment and control groups combined reported that it was "only" or "mostly" the job of the preschool to teach math and reading to preschoolers and there were no differences

16. In a pretest we asked separately about math and reading. Because parents almost always gave the same answer for math as they did for reading we combined them in the survey for the whole PACT sample.

between the treatment and control group on this measure.

In the same survey we also asked, "Think about a child's ability to talk and use language. How much do you think this ability is set from birth and how much is based on what parents teach the child?" If the information provided on the tablets is an important part of the treatment effect we would expect that parents in the treatment group would be less likely to answer that they believe ability is set. Only 6% of the parents indicated that child ability was set all or mostly at birth and there was again no difference between the treatment and control group.

The survey results suggest that parents in both the treatment and control groups believe that spending educational time with their child will increase their child's skills and that it is the parents' job to do this. A fuller description of these results can be found in the appendix. It is perhaps not surprising that there is so little difference in the responses to these questions given that an important mission of Head Start is to convey to parents the importance of engaging their children and especially the importance of reading to their children. Parents clearly hear the message but many do not act on the information that they receive. From this we conclude that additional information about reading and parental engagement as presented in the PACT intervention played almost no role in the treatment effect.

1.5 Robustness Checks

To test the robustness of our estimates we estimate several different specifications for our main outcome of interest. In the appendix, we estimate the same specifications for additional outcomes.

1.5.1 Drop-Outs

Of the 169 parents, eleven returned their tablets and stopped their participation in PACT before the end of the intervention. Six of these parents belonged to the control group and all of them stopped during the first week of the intervention. Two of them recorded some reading. Of the five treated parents who stopped, three did so during the first week and two did so during the second week. We asked all eleven parents why they dropped out. The reasons included that they did not enjoy the check-ins (when we downloaded data from the tablets); that the child was too young to use the tablet; that the child was not interested in reading on a tablet; that the children in the home were fighting over the tablet; that the intervention felt like an invasion of privacy; and that the child was sick.

The results in Table 1.2 include the parents who dropped out and count them as not spending any time on the app after they dropped out. If we omit the eleven parents who dropped out, the treatment effect on the number of minutes spent reading with the app increases from 88.3 to 92.3 minutes.

1.5.2 Outliers

Figure 1.2 shows the distribution of reading time in 10-minute intervals, by treatment status. It shows that the control group's average total minutes of reading over the six weeks is clustered at the lower end of reading time. It also shows that for both the control and treatment groups there are individual parents who are on the far right tail of the distribution. Table 1.4 shows the sensitivity of our results on reading time to the omission of outliers. Column 1 shows our estimation for the effect of the treatment on reading time for the full sample. Columns 2, 3 and 4 exclude parents who read more than 600 minutes (three standard deviations (SD) above the treated group mean), 450 minutes (two SD) and 300 minutes (one SD), respectively.[17] Not surprisingly the coefficient on the treatment declines as we omit outliers. However, the effect sizes remain very large, ranging from 1.38 in the second column to 1.00 in the fourth column.

We also tested the sensitivity of the results to omitting those who did not read at all, as shown in Column 5 in Table 1.4. Among parents who read at all, parents in the treatment

17. The standard deviation for the treated group is 148 minutes.

group read 83.4 more minutes than those in the control group, which equals an effect size of .92.

1.5.3 Covariates, School Fixed Effects and Clustering

Of the 169 parents, 160 answered our initial survey. From those 160 parents, 151 have complete information (i.e., non-missing values) for the set of covariates described in Table 1.1 (except family income and race and ethnicity because these had a large number of missing data). The 18 parents with some missing information include 10 in the treatment group and 8 in the control group. The first column in Table 1.5 shows the baseline estimation for the 169 parents (from Table 1.2). The second column shows the results for the sample with complete information in the set of covariates in Table 1.1. The coefficient on treatment drops from 88.3 to 79.5. Column 3 adds school fixed effects, column 4 controls the covariates from Table 1.1 and column 5 adds school clusters. The results in column 5 show that the treatment effect and its statistical significance remains virtually unchanged when we condition on a large set of covariates, use school fixed effects and when we cluster on schools.

1.6 Conclusions

The PACT intervention increased the amount of time that parents in the treatment group read to their children by over 88 minutes over six weeks. We speculated that one reason that disadvantaged parents do not read more to their children is that they discount the future, resulting in impatience. Our evidence that less patient parents read the least to their children but that the intervention had the greatest impact on less patient parents supports this hypothesis.

A primary motivation for this intervention was the fact that programs to help parents engage with their children have largely been unsuccessful. Consequently we were mainly interested in motivating parents to engage in the specific task of using the app to read and not necessarily in increasing parents' total reading time. However, increasing total reading time is also an important goal and as such it raises the question of whether parents substituted reading on the tablet for reading non-electronic books (or other forms of positive engagement with their children). Of course the question of substitution arises in many RCTs because we seldom can measure all the alternative behaviors that participants might have engaged in absent the intervention. But in the case of PACT, parents and children might have found the tablets new and fun and the number and variety of books on the tablet may also have made substitution of the tablet for paper books likely. We have no direct evidence of the extent of such substitution.

However, some evidence suggests that little substitution occurred overall or at least that it did not occur differentially for treatment and control group parents. First, the amount of time that the parents in our sample spend reading to their children is very small. As Table 2 shows the control group read only 63.3 minutes over the entire six-week intervention. This equals about 1.5 minutes per day of reading. Parents in the control group all had the tablet with the reading app, so if these were enticing, the amount of reading absent the behavioral tools must be close to zero leaving little room for substitution. We have no reason to believe that the pre-intervention reading time of the treatment group was any greater than the control group's reading time during the intervention. Thus it does not seem unreasonable to assume that prior to the intervention the amount of time that parents who participated in PACT read to their children was extremely small and therefore that there was little room for substitution.

Assuming little or no substitution we can put the results from PACT in perspective using the results from the related literature. For example, as we described above Price (2010) estimates that an extra half hour of parents' reading to a child per week would raise the children's reading scores by about 5% of a standard deviation per year. If that increase were cumulative, after 10 years the increase would be half a standard deviation. However, Price's sample is all mothers in the NLSY so the sample probably has a higher mean for reading time than our sample of disadvantaged parents. If we assume that the relationship between reading time and test scores is concave downward, then the effect of the PACT treatment on test scores of disadvantaged children would be greater than is suggested by Price's results. In our intervention the treatment group read on average about five stories per week, compared to about two stories read by the parents in the control group. Kalbs and von Ours (2014) find that reading 3-5 times a week to children (aged 4-5) is associated with about a half a standard deviation increase in a cognitive skill index compared to reading less than three times per week. Their sample is also less disadvantaged than the PACT sample so the effect of PACT on disadvantaged children is likely to be greater. Finally, it is worth noting that several papers find that reading to children not only increases literacy skill but also numeracy and non-cognitive skills. If so the effect of PACT on the amount of time that parents spending reading to their children may have even more far-reaching benefits than suggested by scores on tests of literacy.

This paper has demonstrated that low-cost goal setting and reminders can motivate parents to do things that they want to do but often fail to do. This is important because many interventions have tried to change parental behavior with little success. Because these behavioral tools can be cost effective, they provide a promising way to help parent engage their children more often and more productively.

1.7 References

Agee, Mark, and Thomas Crocker. 1996. "Parents' Discount Rates for Child Quality. " Southern Economic Journal 63: 36-50.

Altintas, Evrim. "Parents' Time with Children: Micro and Macro Perspectives." PhD diss., Nuffield College, University of Oxford, UK, 2012.

Andreoni, James, and Charles Sprenger. 2012. "Estimating Time Preferences from Convex Budgets." American Economic Review 102(7): 33333356.

Andreoni, James, Michael A. Kuhn, and Charles Sprenger. 2013. "On Measuring Time Preferences." NBER Working Paper 19392, National Bureau of Economic Research.

Ashraf, Nava. 2013. "Rx: Human Nature: How Behavioral Economics is Promoting Better Health Around the World." Harvard Business Review 91(4).

Baker, Amy J.L., Chaya S. Piotrkowski, and Jeanne Brooks-Gunn. 1999. "The Home Instructional Program for Preschool Youngsters (HIPPY)." The Future of Children 9(1): 116-133.

Bnabou, Roland, and Jean Tirole. 2003. "Intrinsic and Extrinsic Motivation." The Review of Economic Studies 70: 489-520.

Bnabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." The American Economic Review 96(5): 1652-1678.

Benjamin, Daniel J., James J. Choi, and A. Joshua Strickland. 2010. "Social Identity and Preferences." American Economic Review 100(4): 1913-28.

Bergman, Peter. 2015. "Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment." CESifo Working Paper 5391.

Besley, Timothy, and Maitreesh Ghatak. 2005. "Competition and Incentives with Motivated Agents." American Economic Review 95(3): 616-636.

Bryan, Gharad, Dean Karlan, and Scott Nelson. 2010. "Commitment Devices." Annual Review of Economics 2(1): 671-698.

Burks, Stephen, Jeffrey Carpenter, Lorenz Gtte, and Aldo Rustichini. 2012. "Which Measures of Time Preference Best Predict Outcomes: Evidence from a Large-Scale Field Experiment." Journal of Economic Behavior and Organization 84 (1): 308-320.

Carneiro, Pedro and James J. Heckman. 2003. "Human Capital Policy." In Inequality in America: What Role for Human Capital Policies?, edited by James J. Heckman, Alan B. Krueger, and Benjamin M. Friedman. Cambridge, MA: MIT Press.

Castillo, Marco, Paul Ferraro, Jeffrey Jordan, and Ragan Petrie. 2011. “The Today and Tomorrow of Kids: Time Preferences and Educational Outcomes of Children.” Journal of Public Economics 95(11): 13771385.

Castleman, Benjamin L., and Lindsay Page. 2013. “Summer Nudging: Can Personalized Text Messages and Peer Mentor Outreach Increase College Going Among Low-Income High School Graduates?” Center for Education Policy and Workforce Competitiveness Working Paper 9, Charlottesville, VA: University of Virginia.

Chabris, Christopher F., David Laibson, Carrie L. Morris, Jonathon P. Schuldt, and Dmitry Taubinsky. 2008. “Individual Laboratory- Measured Discount Rates Predict Field Behavior.” Journal of Risk and Uncertainty 37(2/3).

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

Cunha, Flavio, Heckman, James J., Lochner, Lance J. and Masterov, Dimitriy V. 2006. “Interpreting the Evidence on Life Cycle Skill Formation,” In Handbook of the Economics of Education, edited by E. Hanushek and F. Welch. Amsterdam: North Holland, Chapter 12, pp. 697-812.

Cuhna, Flavio, Jennifer Culhane, and Irma Elo. 2013. “Eliciting Maternal Expectations about the Technology of Cognitive Skill Formation.” NBER Working Paper 19144. National Bureau of Economic Research.

Dixit, Avinash. 2002. “Incentives and Organizations in the Public Sector: An Interpretative Review.” Journal of Human Resources 37(4): 696-727.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2010. “Are Risk Aversion and Impatience Related to Cognitive Ability?” American Economic Review 100(3): 1238 1260.

Eckel, Catherine, Cathleen Johnson, and Claude Montmarquette. 2005. “Saving Decisions of the Working Poor, Short- and Long-Term Horizons.” In Field Experiments in Economics, Research in Experimental Economics, edited by Jeffrey P. Carpenter, Glenn W. Harrison, and John A. List. Greenwich, CT: JAI Press.

Fiorini, Mario, and Michael P. Keane. 2014. “How the Allocation of Children’s Time Affects Cognitive and Noncognitive Development.” Journal of Labor Economics 32(4): 787-836.

Frederick, Shane, George Loewenstein, and Ted O’Donoghue. 2002. “Time Discounting and Time Preference: A Critical Review.” Journal of Economic Literature 40: 351401.

Galvan, Adrianna. 2012. ”Risky Behavior in Adolescents: The Role of the Developing Brain.” In The Adolescent Brain: Learning, Reasoning, and Decision Making, edited by Valerie F. Reyna, Sandra B. Chapman, Michael R. Dougherty and Jere Confrey. Washington, D.C.: American Psychological Association.

Gennetian, Lisa A., and Eldar Shafir. 2015. “The Persistence of Poverty in the Context of Financial Instability: A Behavioral Perspective.” Journal of Policy Analysis and Management 34(4): 904-936.

Gilovich, Thomas, Dale Griffin, and Daniel Kahneman, D. (Eds.). 2002. Heuristics and Biases: The Psychology of Intuitive Judgment. New York: Cambridge University Press.

Gin, Xavier, Dean Karlan, and Jonathan Zinman. 2010. “Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation.” American Economic Journal: Applied Economics 2(4): 213-235.

Gneezy, Uri, and Aldo Rustichini. 2000. “Pay Enough or Don't Pay at All.” Quarterly Journal of Economics 115(3): 791-810.

Golsteyn, Bart, Hans Grnqvist, and Lena Lindahl. 2013. “Time Preferences and Lifetime Outcomes.” IZA Discussion Paper 7165. The Institute for the Study of Labor (IZA) in Bonn.

Gomby, Deanna S. 1999. “Understanding Evaluations of Home Visitation Programs.” Future of Children 9 (1): 2743.

Guryan, Jonathan, Erik Hurst, and Melissa Schettini Kearney. 2008. “Parental Education and Parental Time with Children.” Journal of Economic Perspectives 22(3): 23-46.

Hale, L., Berger, L. M., LeBourgeois, M. K., and J. Brooks-Gunn. 2011. A longitudinal study of preschoolers' language-based bedtime routines, sleep duration, and well-being. Journal of Family Psychology, 25, 423433.

Harris, Yvette R., Denise Terrel, and Gordon Allen. 1999. “The Influence of Education Context and Beliefs on the Teaching Behavior of African American Mothers.” Journal of Black Psychology 25(4): 490-503.

Heckman, James J., and Dimitriy V. Masterov. 2004. “The Productivity Argument for Investing in Young Children.” Technical Report Working Paper 5. Washington, DC: Committee on Economic Development.

Heckman, James J. 2008. “Schools, Skills and Synapses,” Economic Inquiry, 46(3): 289-324.

Heckman, James J., and Tim Kautz. 2012. ”Hard Evidence on Soft Skills.” Labour Economics 19(4): 451-64.

Hurst, Erik. 2010. Brookings Papers on Economic Activity. Comments and discussion. Washington, DC: The Brookings Institution.

Kalil, Ariel, Rebecca M. Ryan, and Michael R. Corey. 2012. “Diverging Destinies: Maternal Education and the Developmental Gradient in Time with Children.” Demography 49(4): 1361-1383.

Kalil, Ariel, Susan E. Mayer, William Delgado, and Lisa Gennetian. “How Parents Feel When They Parent: Explaining Education Based Differences in Time Spent with Children.” Unpublished paper, Harris School of Public Policy, University of Chicago, Chicago, IL, 2015.

Knudsen, Eric I., James J. Heckman, Judy Cameron, and Jack P. Shonkoff. 2006. ”Building America’s Future Workforce: Economic, Neurobiological and Behavioral Perspectives on Investment in Human Skill Development.” Proceedings of the National Academy of Sciences 103(27): 10155-10162.

Kalb, Guyonne and Jan C. van Ours. 2014. “Reading to young children: A head-start in life?” Economics of Education Review 40, 124.

Kloosterman, R., Notten, N., Tolsma, J., and G. Kraaykamp. 2011. “The effects of parental reading socialization and early school involvement on children’s academic performance: A panel study of primary school pupils in the Netherlands. European Sociological Review, 27, 291306.

Lavecchia, Adam M., Heidi Liu, and Philip Oreopoulos. 2014. ”Behavioral Economics of Education: Progress and Possibilities.” NBER Working Paper 20609. National Bureau of Economic Research.

Lawrance, Emily C. 1991. “Poverty and the Rate of Time Preference: Evidence from Panel Data.” The Journal of Political Economy 99 (1), 5477.

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao. 2013. “Poverty Impedes Cognitive Function.” Science 30: 976-980.

Meier, Stephan, and Charles Sprenger. 2010. “Present-Biased Preferences and Credit Card Borrowing.” American Economic Journal - Applied Economics 2(1):193210.

Mol, S. E., and A.G. Bus. 2011. To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. Psychological Bulletin, 137, 267296.

Noel, Amber, Patrick Stark, and Jeremy Redford. 2015. “Parent and family involvement in education, from the National Household Education Surveys program of 2012: First look.” (NCES 2013- 028.REV). Washington, DC: U.S. Department of Education.

Olds, David, Charles Henderson, and Harriet Kitzman. 1994. “Does Prenatal and Infancy Nurse Home Visitation Have Enduring Effects on Qualities of Parental Caregiving and Child Health at 25 to 50 Months of Life?” Pediatrics 93: 8998.

Pabilonia, Sabrina W., and Younghwan Song. 2013. “Single Mothers’ Time Preference, Smoking, and Enriching Childcare: Evidence from Time Diaries.” Eastern Economic Journal 39(2): 227-255.

Price, Joseph. 2010. “The Effect of Parental Time Investments: Evidence from Natural Within-Family Variation.” Brigham Young University and National Bureau of Economic Research.

Raikes, H., Pan, B. A., Luze, G., Tamis-LeMonda, C. S., Brooks-Gunn, J., Constantine, J., Tarullo, L. B., Raikes, H. A., and Rodriguez, E. T. 2006. "Mother-child bookreading in low-income families: Correlates and outcomes during the first three years of life." Child Development, 77, 924953.

Ramey, Gary, and Valerie A. Ramey. 2010. "The Rug Rat Race." Brookings Papers on Economic Activity 129-176.

Richburg-Hayes, Lashawn, Caitlin Anzelone, Nadine Dechausay, Saugato Datta, Alexandra Fiorillo, Louis Potok, Matthew Darling, and John Balz. 2014. "Behavioral economics and social policy: Designing innovative solutions for programs supported by the Administration for Children and Families." (OPRE Report N. 2014-16a). Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Rick, Scott, and George Loewenstein. 2008. "The Role of Emotion of Economic Behavior." In Handbook of Emotions, edited by Michael Lewis, Jeannette Haviland-Jones, and Lisa Feldman Barrett. New York: Guilford Publications.

Roggman, Lori, Lisa Boyce, and Mark S. Innocenti. 2008. "Developmental Parenting: A Guide for Early Childhood Practitioners." Baltimore: Paul H. Brookes Publishing.

Schweinhart, Lawrence, Jeanne Montie, Zongping Xiang, W. Steven Barnett, Clive R. Belfield, and Milagros Nores. 2005. Lifetime Effects: The HighScope Perry Preschool Study Through age 40. Ypsilanti, MI: High/Scope Press.

Spears, Dean. 2011. "Economic Decision-Making in Poverty Depletes Behavioral Control." The B. E. Journal of Economic Analysis and Policy 11(1) .

Stanovich, Keith E., Richard F. West, and Maggie E. Toplak. 2012. "Judgment and Decision Making in Adolescence: Separating Intelligence From Rationality." In The Adolescent Brain: Learning, Reasoning and Decision Making, edited by Valerie F. Reyna, Sandra B. Chapman, Michael R. Dougherty, and Jere Confrey. Washington, DC: American Psychological Association.

Sutter, Matthias, Martin G. Kocher, Daniela Glaetzle-Ruetzler, and Stefan T. Trautmann. 2013. "Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior." American Economic Review 103(1):510531.

Thaler, Richard. 2015. Misbehaving: The Making of Behavioral Economics. New York: W.W. Norton and Company, Inc.

Villena-Rodn, Benjamn, and Cecilia Ros-Aguilar. 2012. "Causal Effects of Maternal Time Investment on Children's Cognitive Outcomes." Center for Applied Economics Working Paper 285.

Wagner, Mary, Donna Spiker, and Margaret Inman Linn. 2002. "The Effectiveness of the Parents As Teachers Program with Low-Income Parents and Children." Topics in Early Childhood Special Education 22(2): 6781.

Waldfogel, Jane, and Elizabeth Washbrook. 2011. “Early Years Policy.” Child Development Research 2011.

York, Benjamin, and Susanna Loeb. 2014. “One Step at a Time: The Effects of an Early Literacy Text Messaging Program for Parents of Preschoolers.” NBER Working Paper 20659. National Bureau of Economic Research.

Zaveri, Heather, Andrew Burwick, and Erin Maher. 2014. Home Visiting: The Potential for Cost Savings from Home Visiting Due to Reductions in Child Maltreatment. Washington, DC: Casey Family Programs.s.), Handbook of the Economics of Education, vol. 4. Elsevier, Amsterdam, pp. 1-181.

1.8 Appendix

1.8.1 Recruitment

We recruited parents at eight subsidized preschool programs in Chicago, Illinois. These centers were selected partly because they represented the diversity of centers in Chicago, partly because they are large, and partly because are located in areas accessible for our research assistants. They were located throughout the city and have a racially and ethnically diverse population as we document in the main text in our paper. Figure 1.3 presents a map of the City of Chicago with the geo-referenced preschools.

Together these preschool programs served about 965 children at the time we began recruiting at the beginning of the 2014-2015 school year. Given that the material we developed for our treatment was in English or Spanish, parents and children who did not speak these languages were not eligible for PACT. Based on conversations with the staff of the preschools we estimate that about 15 percent of the 965 children were not eligible due to the language condition. We estimate that an additional 15 percent were siblings of another child in the center and only one sibling could be a focal child in PACT. Therefore, across the 8 centers, about 580 children were eligible to participate in PACT.

We spent between 6-10 hours in each center inviting parents to participate when they dropped off or picked up their children. Of the estimated 580 children whose parents were eligible, 313 signed up on site and we could contact 220 of these. The main reason that we could not contact parents was that the phone number we had was not accurate. Among this group, 51 parents declined to participate. The main reason for not participating was not having time to complete the orientation session, which we describe below. In the end 169 parents agreed to participate in the PACT intervention.

1.8.2 Orientation Sessions and Randomization

All participating parents were given an immediate on-site orientation session, that took about 30 minutes on average. Prior to the orientation parents were randomly assigned to the treatment of the control group. Once parents agreed to participate in PACT they were assigned an electronic tablet to borrow for the six-week intervention. Each tablet had a numeric identifier, known by the research team only. Parents assigned tablets with even identifiers were assigned to the treatment group (N=84), while odd identifiers were assigned to the control group (N=85). Tablets for the treatment group were assigned to every other parent who agreed to participate.

During the orientation parents signed all necessary forms and were instructed in the use of the tablet and the reading application.[18] During the orientation parents in both the control and treatment groups were asked to watch six short instructional videos. Two of these were on how to use the reading application and four were an introduction to the PACT study. Parents were required to sign a consent form, and a pledge to borrow, safeguard, and return the tablet at the end of the program.

While under the supervision of our research team, all parents were given the opportunity to try the digital reading application on their own with a hard copy set of directions that they followed at their own pace. In doing so, they made a video of themselves using the reading application on the tablet, mimicking what they would do when reading to their child.

At the end of the orientation sessions, all parents took a short survey on the tablet. This was the first of three surveys that we administered during the intervention, as we explain in the next section.

18. The rest of the functionalities of the tablet like Internet access or additional applications were disabled.

1.8.3 Survey Response

We collected data in 3 surveys. Survey 1, administered during the orientation session, asked parents to report characteristics of the child, the household, and themselves. Survey 2 was a time preference assessment that we discuss below. Survey 3 was administered at the end of the six-week intervention. It asked questions about parenting beliefs and practices. Table 1.6 show the number of parents who answered each survey, by treatment status. Almost all parents completed survey 1 (95%), while two thirds answered surveys 2 and 3.

1.8.4 Representativeness of PACT participants

Overall the parents participating in the PACT intervention are similar to a national sample of parents of children in Head Start programs on characteristics collected in the Head Start Family and Child Experiences Survey (FACES) survey except that a higher proportion of PACT families are Spanish speaking and Hispanic, which is not surprising because language was an eligibility criteria for participation in PACT. Table 1.7 reports the comparison of the PACT participants to the parents of children entering Head Start on 2010, the latest year for which FACES data is available.

1.8.5 Time Preferences

Our main hypothesis about why disadvantaged parents spend less educational time with their children is that disadvantaged parents are prone to impatience or that they have a high discount rate for the future. We examine if parents in our sample who are more future-oriented read more and respond to the behavioral tools more than parents in our sample who are more present oriented. We next describe the Time Preferences Task, explain our estimation procedures (for our parameters of interest and their associated standard errors), and present the time preferences questionnaire.

1.8.5.A. Time Preferences Task

Following Andreoni and Sprenger (2012, 2013), we designed a survey to measure time preferences. This survey corresponds to the PACT survey 2, which was administered on electronic tablets after random assignment to parents in both the treatment and control groups. Parents were given a set of instructions and an example before they began the task. They were given the opportunity to ask questions before the task and during it. Instructions and an example were also available in an audio recording on the tablet that parents could play at any time while they performed the task. All the parents in a particular center were given the time preference task on the same day. To induce parents to answer truthfully, all respondents were entered into a lottery and one parent was randomly selected to actually receive the payout. For that parent the actual amount and time for the payout was determined by randomly selecting one of the 15 payouts that the parent preferred. Parents always had a greater than one in ten chance of receiving a payout and they we informed of these odds prior to taking participating in the task. Parents were asked to choose between an amount of money that they could receive immediately or an amount that they could receive later. In total, each parent answered 15 of these questions. In an effort to simplify the choice situation, each of the questions offered four choices only (two inner solutions and two corner solutions), as in Andreoni et al (2013). The first five questions offered payments immediately and three weeks later, the next five questions offered payments immediately and six weeks later, and the last five questions offered payments in three and six weeks. Within each time horizon, each question presented an increasing price for the earlier payment. The set of choices is presented below. The 3 and 6- week time horizons were chosen because the intervention lasted six weeks so the payout could be made during or at the end of the intervention.

1.8.5.B. Time Preferences Estimation

We use the Convex Time Budget (CTB)[19] method (first introduced by Andreoni and Sprenger, 2012) to estimate both a discount rate for parents and their present bias, which have shown to best predict outcomes in previous research (Burks et al. 2012). We follow closely Andreoni et al (2013) and assume that parents have preferences for the experimental payments X in period t and t+k can be represented by a utility function U(Xt , Xt+k). U is a time-separable, constant relative risk averse function with a quasi-hyperbolic structure for discounting (Laibson, 1997; O'Donoghue and Rabin, 1999), given by

$$U(X_t, X_{t+k}) = {X_t}^{\alpha} + \beta^{t_0}\delta^{k}{X_{t+k}}^{\alpha}$$

that parents maximize subject to a budget constraint given by,

$$PX_t + X_{t+k} = Y$$

where, t_0 is an indicator function that takes value 1 if time t is in the present (now); β denotes the extent of present bias ; δ denotes the discount rate ; α denotes the curvature of the utility function. ; P denotes the price ratio or gross interest rate; and Y denotes the total experimental budget. The first order conditions from maximizing the utility function subject to the budget constraint yield the standard condition,

$$\frac{MgU_{X_t}}{MgU_{X_{t+k}}} = P$$

indicating that the ratio of the marginal utilities from the earlier payment and later

19. The CTB method identifies the curvature in the utility function over time by estimating a respondent's sensitivity to changing interest rates. The key in using this approach is to vary the implicit interest rate in the options presented across subsequent sets of options. The sensitivity to changing interest rates across the question sequences identifies the utility curvature; the time preference is identified through the stated preference over the timing of payments.

payments should equalize the price ratio P. This implies that

$$\frac{X_t}{X_{t+k}} = \left(P\beta^{t_0}\delta^k\right)^{\frac{1}{\alpha-1}}$$

Taking logs,

$$ln\left(\frac{X_t}{X_{t+k}}\right) = \underbrace{\left(\frac{ln(\beta)}{\alpha-1}\right)}_{\pi_0} t_0 + \underbrace{\left(\frac{ln(\delta)}{\alpha-1}\right)}_{\pi_1} k + \underbrace{\left(\frac{1}{\alpha-1}\right)}_{\pi_2} ln(P)$$

Assuming an additive error structure, the estimation from the equation above pprovides the estimated reduced form parameters $\widehat{\pi_0}, \widehat{\pi_1}$ and $\widehat{\pi_2}$, that allow us to identify our structural parameters $\widehat{\alpha}, \widehat{\beta}, \widehat{\delta}$.

We compute the associated standard errors using the delta method. From the equation, it can be ascertained that the variation in P identifies α; variation in k identifies ; and variation in t0 identifies β. The interpretation of the structural parameters is as follows. If $\alpha = 1$ the utility function is linear. If $\alpha < 1$ individuals have convex preferences in (sooner, later) space. More curvature implies that individuals prefer interior solutions (they prefer to spread payments) over corner solutions. Lower δ implies less weight to the future, or more patience. If $\beta = 1$ there is no present bias. If $\beta < 1$ there is present bias.

1.8.5.C. Recovering Standard Errors using the Delta Method

We start with α, because it is a function of only one reduced form parameter

$$\alpha = g(\pi_2) = 1 + \frac{1}{\pi_2}$$

We need to assume that $\widehat{\pi_2}$ is a consistent estimator of π_2 and that a CLT can be applied (iid, finite variance). Then,

$$\sqrt{n}(\pi_2 - \widehat{\pi_2}) \xrightarrow{d} N(0, \sigma_2^2) \tag{*}$$

which we will use to get,

$$\sqrt{n}(\alpha - \widehat{\alpha}) \xrightarrow{d} N(0, \sigma_\alpha^2)$$

A first order approximation of $g(\pi_2)$ for π_2 near $\widehat{\pi_2}$,

$$\begin{aligned} g(\pi_2) &\approx g(\widehat{\pi_2}) + g'(\widehat{\pi_2})(\pi_2 - \widehat{\pi_2}) \\ \sqrt{n}(g(\pi_2) - g(\widehat{\pi_2})) &\approx g'(\widehat{\pi_2})\sqrt{n}(\pi_2 - \widehat{\pi_2}) \\ \sqrt{n}(\alpha - \widehat{\alpha}) &\approx g'(\widehat{\pi_2})\sqrt{n}(\pi_2 - \widehat{\pi_2}) \end{aligned}$$

Using (1) and the Slutsky Theorem,

$$\sqrt{n}(\alpha - \widehat{\alpha}) \approx \underbrace{g'(\widehat{\pi_2}) \underbrace{\sqrt{n}(\pi_2 - \widehat{\pi_2})}_{\xrightarrow{d} N(0,\sigma_2^2)}}_{\xrightarrow{d} N(0,[g'(\widehat{\pi_2})]^2\sigma_2^2)}$$

Therefore

$$\sqrt{n}(\alpha - \widehat{\alpha}) \xrightarrow{d} N(0, [g'(\widehat{\pi_2})]^2\sigma_2^2)$$

Hence an estimator of $Var(\alpha)$ is given by $\widehat{\sigma_\alpha^2} = [g'(\widehat{\pi_2})]^2\widehat{\sigma_2^2}$

An analogous procedure can be applied for $\beta = f(\pi_0, \pi_2) = exp(\pi_0/\pi_2)$ and $\delta = f(\pi_1, \pi_2) = exp(\pi_1/\pi_2)$

Which delivers

$\sigma_\beta^2 = f_{\pi_0}^2\sigma_0^2 + 2f_{\pi_0}f_{\pi_2}\sigma_{02} + f_{\pi_2}^2\sigma_2^2 = (f_{\pi_0}\sigma_0 + f_{\pi_2}\sigma_2)^2$

$\sigma_\delta^2 = f_{\pi_1}^2\sigma_1^2 + 2f_{\pi_1}f_{\pi_2}\sigma_{12} + f_{\pi_2}^2\sigma_2^2 = (f_{\pi_1}\sigma_1 + f_{\pi_2}\sigma_2)^2$

where $f_{\pi_i} = \frac{\partial f(\cdot)}{\partial \pi_i}$ for i=0,1,2.

Therefore the standard errors are given by,

- $SE(\widehat{\alpha}) = g'(\widehat{\pi_2})\widehat{\sigma_2}$
- $SE(\widehat{\beta}) = f_{\pi_0}\widehat{\sigma_0} + f_{\pi_2}\widehat{\sigma_2}$
- $SE(\widehat{\delta}) = f_{\pi_1}\widehat{\sigma_1} + f_{\pi_2}\widehat{\sigma_2}$

1.8.5.D. Time Preferences Questionnaire

Payout Question 1 of 15

Look at the four ways you could get money if you win. You get some money this week, and the rest in three weeks.

Click on the button next to the choice you like best.

◉ **Choice A**
$24 guaranteed this week and $0 in 3 weeks

○ **Choice B**
$16 guaranteed this week and $8 guaranteed in 3 weeks

○ **Choice C**
$8 guaranteed this week and $16 guaranteed in 3 weeks

○ **Choice D**
$0 this week and $24 guaranteed in 3 weeks

Payout Question 2 of 15

Each page has four different ways you could get your money if you win.

Look at the choices and click on the button next to the one choice you like best.

◉ **Choice A**
$21 guaranteed this week and $0 in 3 weeks

○ **Choice B**
$14 guaranteed this week and $8 guaranteed in 3 weeks

○ **Choice C**
$7 guaranteed this week and $16 guaranteed in 3 weeks

○ **Choice D**
$0 this week and $24 guaranteed in 3 weeks

Payout Question 3 of 15

Here are four different ways to get the money if you win.

Click on the button next to the one choice you like best.

◉ **Choice A**
$18 guaranteed this week and $0 in 3 weeks

○ **Choice B**
$12 guaranteed this week and $8 guaranteed in 3 weeks

○ **Choice C**
$6 guaranteed this week and $16 guaranteed in 3 weeks

○ **Choice D**
$0 this week and $24 guaranteed in 3 weeks

Payout Question 4 of 15

Which of these do you like better?

- **Choice A**
  $15 guaranteed this week and $0 in 3 weeks
- **Choice B**
  $10 guaranteed this week and $8 guaranteed in 3 weeks
- **Choice C**
  $5 guaranteed this week and $16 guaranteed in 3 weeks
- **Choice D**
  $0 this week and $24 guaranteed in 3 weeks

Payout Question 5 of 15

If you win, which way would you like to be paid? Choose one by clicking next to it.

- **Choice A**
  $12 guaranteed this week and $0 in 3 weeks
- **Choice B**
  $8 guaranteed this week and $8 guaranteed in 3 weeks
- **Choice C**
  $4 guaranteed this week and $16 guaranteed in 3 weeks
- **Choice D**
  $0 this week and $24 guaranteed in 3 weeks

Payout Question 6 of 15

Please Note: Now the choices include money this week and in *six* weeks. Pick the one you like most.

- **Choice A**
  $24 guaranteed this week and $0 in 6 weeks
- **Choice B**
  $16 guaranteed this week and $8 guaranteed in 6 weeks
- **Choice C**
  $8 guaranteed this week and $16 guaranteed in 6 weeks
- **Choice D**
  $0 this week and $24 guaranteed in 6 weeks

Payout Question 7 of 15

This is a new set of choices but the payouts are still this week and in six weeks. Please choose one.

- ○ **Choice A**
  $21 guaranteed this week and $0 in 6 weeks
- ○ **Choice B**
  $14 guaranteed this week and $8 guaranteed in 6 weeks
- ○ **Choice C**
  $7 guaranteed this week and $16 guaranteed in 6 weeks
- ○ **Choice D**
  $0 this week and $24 guaranteed in 6 weeks

Payout Question 8 of 15

Which one of these choices do you prefer? Remember that if you win, you will get your money on one of the schedules that you pick in this survey.

- ◉ **Choice A**
  $18 guaranteed this week and $0 in 6 weeks
- ○ **Choice B**
  $12 guaranteed this week and $8 guaranteed in 6 weeks
- ○ **Choice C**
  $6 guaranteed this week and $16 guaranteed in 6 weeks
- ○ **Choice D**
  $0 this week and $24 guaranteed in 6 weeks

Payout Question 9 of 15

Now four different choices. Which one of these do you like best?

- ○ **Choice A**
  $15 guaranteed this week and $0 in 6 weeks
- ○ **Choice B**
  $10 guaranteed this week and $8 guaranteed in 6 weeks
- ○ **Choice C**
  $5 guaranteed this week and $16 guaranteed in 6 weeks
- ○ **Choice D**
  $0 this week and $24 guaranteed in 6 weeks

Payout Question 10 of 15

Please select the payout you like best of these four:

- (selected) **Choice A**
  $12 guaranteed this week and $0 in 6 weeks
- **Choice B**
  $8 guaranteed this week and $8 guaranteed in 6 weeks
- **Choice C**
  $4 guaranteed this week and $16 guaranteed in 6 weeks
- **Choice D**
  $0 this week and $24 guaranteed in 6 weeks

Payout Question 11 of 15

Please note: now the money comes partly in *three* weeks and partly in *six* weeks. Pick the one you prefer.

- (selected) **Choice A**
  $24 guaranteed in 3 weeks and $0 in 6 weeks
- **Choice B**
  $16 guaranteed in 3 weeks and $8 guaranteed in 6 weeks
- **Choice C**
  $8 guaranteed in 3 weeks and $16 guaranteed in 6 weeks
- **Choice D**
  $0 in 3 weeks and $24 guaranteed in 6 weeks

Payout Question 12 of 15

The choices continue to combine payments in *three weeks* and in *six weeks*. Click a button beside the one you like best.

- (selected) **Choice A**
  $21 guaranteed in 3 weeks and $0 in 6 weeks
- **Choice B**
  $14 guaranteed in 3 weeks and $8 guaranteed in 6 weeks
- **Choice C**
  $7 guaranteed in 3 weeks and $16 guaranteed in 6 weeks
- **Choice D**
  $0 in 3 weeks and $24 guaranteed in 6 weeks

Payout Question 13 of 15

Pick the one you like best from these choices:

- (◉) **Choice A**
  $18 guaranteed in 3 weeks and $0 in 6 weeks
- ( ) **Choice B**
  $12 guaranteed in 3 weeks and $8 guaranteed in 6 weeks
- ( ) **Choice C**
  $6 guaranteed in 3 weeks and $16 guaranteed in 6 weeks
- ( ) **Choice D**
  $0 in 3 weeks and $24 guaranteed in 6 weeks

Payout Question 14 of 15

If you win, which of these payment schedules would you pick?

- (◉) **Choice A**
  $15 guaranteed in 3 weeks and $0 in 6 weeks
- ( ) **Choice B**
  $10 guaranteed in 3 weeks and $8 guaranteed in 6 weeks
- ( ) **Choice C**
  $5 guaranteed in 3 weeks and $16 guaranteed in 6 weeks
- ( ) **Choice D**
  $0 in 3 weeks and $24 guaranteed in 6 weeks

Payout Question 15 of 15

This is your last set of choices! Please select the one you like best.

- (◉) **Choice A**
  $12 guaranteed in 3 weeks and $0 in 6 weeks
- ( ) **Choice B**
  $8 guaranteed in 3 weeks and $8 guaranteed in 6 weeks
- ( ) **Choice C**
  $4 guaranteed in 3 weeks and $16 guaranteed in 6 weeks
- ( ) **Choice D**
  $0 in 3 weeks and $24 guaranteed in 6 weeks

1.8.6 Robustness Checks

To test the robustness of our estimates we estimate several different specifications for three outcomes of interest: reading time, number of books, and the fraction of parents who did any reading. In particular, Table 1.8 presents the results for participants by survey response; Table 1.9 shows results excluding outliers and Table 1.10 tests the robustness of our estimates when we include covariates, preschool fixed effects and clustering to our regressions. The treatment impact remains qualitatively the same across our different robustness checks. In addition, the next subsection tests whether our results vary in relation to the parental time preferences.

Time Preferences Sensitivity

From our 169 PACT parents, 112 parents answered the time orientation task. Of the 84 treated parents, 59 participated in the time preference task (70%). Of the 85 control group parents, 53 participated (62%). The difference of 8 percentage points is not significant at 5% (nor at 10%), as shown in Panel ‘ from Table 1.11.

Of the 112 parents who took the time orientation task, 28 gave the same answer to all the 15 time orientation questions, leaving us with 84 parents for whom we could compute a discount rate. Conditional on answering the survey (N=112), there is no significant difference whether response had variation in their answers by treatment status, as shown in Panel 2 from Table 1.11.

We computed the time discounting (delta) parameter for the 84 parents with variation in their answers to the time orientation task. We label this sample “With Time Preferences A”. From the 84 parents, we distinguish:

- “Less Patient Parents A” are those whose score is below the median. (N=42)
- “More Patient Parents A” are those whose score is above the median. (N=42)

This measure of patience is not correlated with treatment status, as presented in Panel 3 from Table 1.11 .

We then classified the 28 parents with no variation in their answers as:

- "Less patient Parents B": if they answered only A or only B.

- "More Patient Parents B": If they answered only C or only D.

This measure of patience is not correlated with treatment status, as presented in Panel 4 from Table 1.11.

Table 1.12 shows the results of estimating the effect of treatment on reading time for the different groups of patient/impatient parents. The effect of the treatment is much stronger for the less patient parents, independent of whether we use measure A or B, as columns (4) and (7) show.

A final test consists in dropping parents that are systematically being inconsistent when making their choices. A choice is classified as consistent if it allocates the same or a larger budget share to the future payment date as a previous choice with a lower gross rate. As the gross rate increases, the law of demand dictates a weakly decreasing allocation to the earlier time point. The number of inconsistent choices at the parent level by treatment status is shown in Table 1.13. We classified a parent to be systematically consistent if the number of inconsistent choices is less or equal to three. This is arbitrary, but the results hold for any other cutoff. Table 1.14 shows that the results hold if we consider the consistent parents only.

1.8.7 Goal Setting and Imputation

Parents in the treatment group were asked to set a goal each week for how much time they would spend reading to their child during the next week. As noted above we included the goal-setting component of the intervention because commitment devices have been shown to change behavior in domains other than parenting. Setting a goal for doing a particular behavior raises the psychological cost of not doing the behavior. Anecdotally we know that parents in PACT attended to the goals that they set. They would often tell members of the research team that they were sorry that they did not meet their goal and express determination to meet the goal. But if not meeting the goal was psychologically costly, then we expect that over the six weeks of the intervention the difference between the goal and the actual amount of reading time would decline.

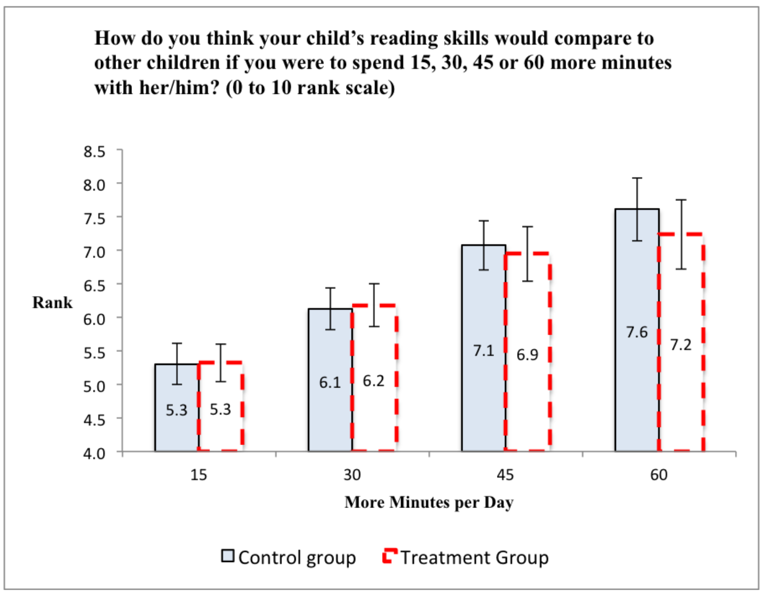
Parents were asked to set a goal when they first received the tablets so for the first week all 84 parents in the treatment group set a goal. Parents were then asked to set a goal at each weekly check in when they dropped off and picked up their tablet at the preschool. If a parent did not bring the tablet in or did not pick it up, we asked the parent to set a goal by text message. However, for a variety of reasons not all parents set a goal every week. By week 2 only 63 parents set a goal by week three the number was 58 and then for the remaining weeks about 50 parents set a goal. Over the six week intervention about 70 percent of parents set a goal for 3 or more of the 6 weeks and about 23 percent set a goal for all six weeks. We assigned the goal of the previous week to the parent that did not set a goal for the following week. If parents missed to set a goal for 2 weeks they are still counted as missing. The assumption here is that parents did not change their goals much over successive weeks. We test this by subtracting the lowest goal from the highest goal for each parent. When we do this we find that about 40% of the parents change their goal in 30 minutes or less, and 60% changed it in an hour or less. After this imputation the number of people reporting a goal declined form 84 in week 1 to 60 in week six.

To look at the difference between the goal and the actual amount of reading we need not only the parent's goal but also the actual amount of reading time. Not all parents brought their tablet in every week so we do not have a weekly reading amount for all parents. But while only 16.5 percent of parents checked in every one of the six weeks, 70 percent checked in for at least 4 of the six weeks. When a parent did not check in for 2 weeks we averaged the total number of minutes at the end of the second week over the previous two weeks. If a parent does not check in for 3 weeks we average the reading time over those three weeks and so on so that everyone who participated in PACT has a reading score for every week.

Using these imputed goals and reading time, we find that the difference between the reading goal and the actual reading time was 102 minutes for week 1. The difference then dropped for the next 2 weeks to 81 minutes and then 74 minutes. Then for the final three weeks the difference stayed between 63 and 64 minutes. This suggests that while parents did set more achievable goals the goals were still se much too optimistically even in the last week the average goal was three times the average actual amount of time the parent spent reading to the child.

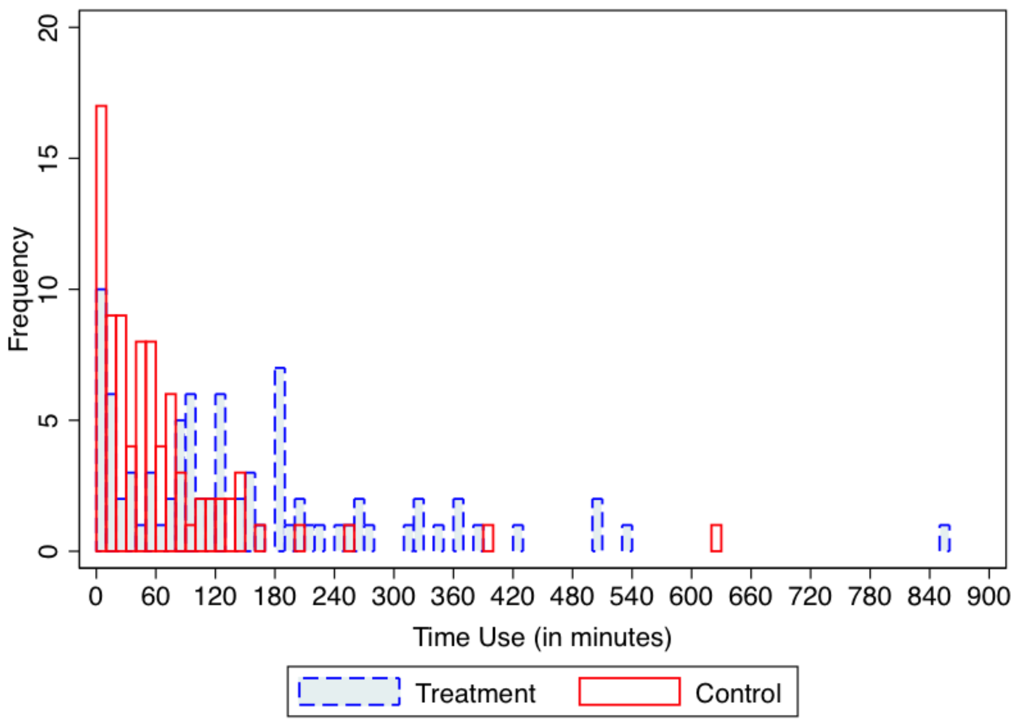
1.9 Figures and Tables

Figure 1.1: Parental expectations: Change in child's rank vs. additional time investment



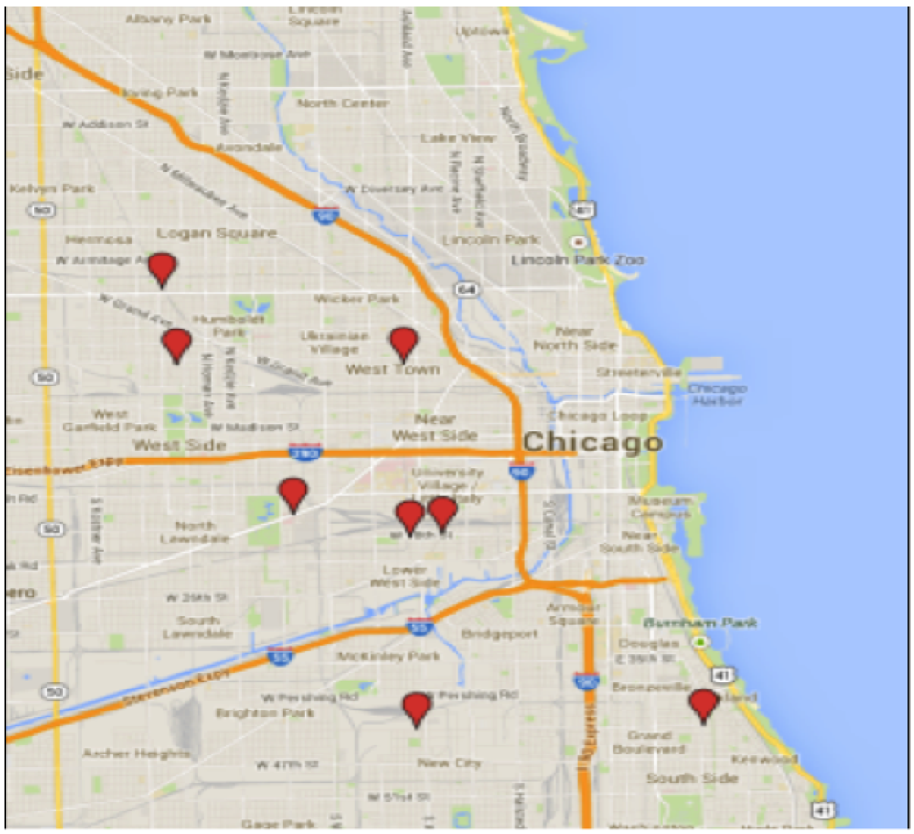
Notes: The data represents answers to the question: "If you were to spend 15, 30, 45 or 60 more minutes per day in educational activities with your child how do you think this would change your child's reading skills compared to all children of the same age across the United States (on a 0-10 scale)?"

Figure 1.2: Frequency distribution of reading minutes by treatment status



Note: The Figure presents the frequency distribution of reading minutes by the control group (red/solid) and the treatment group (light blue/dashed), grouped into 10-minute intervals.

Figure 1.3: PACT Preschool Centers in the Chicago Area



Note: Figure 1.3 presents a map of the City of Chicago with the geo-referenced preschools (depicted by pins). We recruited parents at these eight subsidized preschool programs in Chicago, Illinois. The centers were selected partly because they represented the diversity of centers in Chicago, partly because they are large, and partly because are located in areas accessible for our research assistants. They were located throughout the city and have a racially and ethnically diverse population as we document in the main text in our paper.

Table 1.1: Summary statistics: Total sample, treatment and control groups

Variable	Total Sample	Treatment	Control	Difference	*P-value*
Mother's characteristics					
Female (%)	0.92	0.91	0.94	-0.03	0.63
Single (No Husband/Wife/Partner) (%)	0.61	0.62	0.6	0.03	0.78
Age (years)	31.13	31.55	30.71	0.96	0.39
Black or African American (%)	0.32	0.29	0.36	-0.03	0.72
Hispanic (%)	0.66	0.71	0.59	0.08	0.32
Less than high school (%)	0.28	0.25	0.3	-0.06	0.36
High school diploma (%)	0.12	0.15	0.09	0.06	0.30
GED (%)	0.06	0.1	0.01	0.09	0.10
Some college (%)	0.29	0.24	0.34	-0.10	0.10
Associate's degree (%)	0.13	0.14	0.12	0.02	0.78
Bachelor or higher (%)	0.13	0.12	0.14	-0.02	0.97
Focal child's characteristics					
Age (years)	3.76	3.74	3.77	-0.02	0.89
Boy (%)	0.44	0.47	0.41	0.07	0.28
Born early (%)	0.16	0.20	0.13	0.08	0.07
Born late (%)	0.13	0.11	0.15	-0.04	0.46
Born on time (%)	0.7	0.69	0.72	-0.04	0.41
Less than 5 pounds at birth (%)	0.24	0.24	0.24	-0.01	0.90
Has disability (%)	0.12	0.13	0.11	0.02	0.71
Household characteristics					
Number of children in the household	1.99	1.94	2.05	-0.12	0.45
Number of adults in the household (besides parent)	1.28	1.38	1.18	0.18	0.28
Household income last year ($1,000s)	20.85	18.82	22.59	-3.42	0.31
English spoken at home (%)	0.65	0.62	0.67	-0.01	0.71
Observations	160	80	80		

Notes: Of the 169 parents, 160 answered our initial survey. From those 160 parents, 151 have complete information (i.e., non-missing values) for the set of covariates described in Table 1.1 (except family income and race and ethnicity because these had a large number of missing data). The 18 parents with some missing information include 10 in the treatment group and 8 in the control group.

Table 1.2: The effect of the PACT treatment on measures of parents' reading using the electronic app

	Dependent Variable:		
	(1)	(2)	(3)
	Minutes Read	Number of books	Any Reading
Treatment (δ)	88.32***	16.66***	0.12**
	(18.79)	(4.50)	(0.05)
Constant (α)	63.34***	14.78***	0.84***
	(9.49)	(3.00)	(0.04)
SD Control Group	87.48	27.68	0.37
Effect Size	1.01	0.60	0.31
Observations	169	169	169

Notes: Robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table shows results for the estimation of equation (1) using "Minutes Read", "Number of Books" and "Any reading" as outcomes, each of which are explained as follows. Column 1: "Minutes read" is the number of minutes parents read to their children using the app. Column 2: "Number of books" is the number of books parents read to their children using the app. Column 3: "Any reading" is the fraction of parents who read at all to their children using the app. SD Control Group is the standard deviation of the control group for the respective dependent variable. The effect size is the ratio of the coefficient for the treatment group to the standard deviation of the control group.

Table 1.3: Treatment effects by level of patience

	Dependent Variable: Reading Time			
	(1)	(2)	(3)	(4)
	Baseline: All PACT Parents	Participated in time preference task	Less patient parents	More patient parents
Treatment (δ)	88.32***	82.12***	124.52***	42.26
	(18.79)	(24.08)	(33.25)	(33.95)
Constant (α)	63.34***	83.61***	68.33***	97.26***
	(9.49)	(14.11)	(12.57)	(24.19)
SD Control Group	87.48	102.76	62.95	128.08
Effect Size	1.01	0.8	1.98	0.33
Observations	169	112	54	58

Notes: Robust standard errors in parentheses. Significance levels:* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table shows results for the estimation of equation (1) using "Minutes Read," as the outcome, on different samples, each of which are explained as follows. Column 1: considers all PACT parents. Column 2: considers parents that participated in the time preference task. Column 3 and 4: considers the sample of less and more patient parents, respectively. More patient (less patient) parents are those whose score on the time orientation task is below (above) the median. SD Control Group is the standard deviation of the control group for the respective dependent variable. The effect size is the ratio of the coefficient for the treatment group to the standard deviation of the control group.

Table 1.4: Treatment effects, excluding outliers

	Dependent Variable: Reading Time				
	(1)	(2)	(3)	(4)	(5)
	Baseline:	Excludes parents who:			
	All PACT Parents	read more than 600 minutes	read more than 450 minutes	read more than 300 minutes	did not read at all
Treatment (δ)	88.32***	86.48***	72.48***	51.30***	83.42***
	(18.79)	(15.57)	(13.74)	(10.75)	(19.77)
Constant (α)	63.34***	56.69***	56.69***	52.67***	75.82***
	(9.49)	(6.85)	(6.85)	(5.62)	(10.76)
SD Control Group	87.48	62.78	62.78	51.16	90.68
Effect Size	1.01	1.38	1.15	1.00	0.92
School Fixed Effects	No	No	No	No	No
Covariates	No	No	No	No	No
Observations	169	167	164	155	151

Notes: Robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table shows results for the estimation of equation (1) using "Minutes Read", as outcome, on different samples, each of which are explained as follows. Column 1 includes all the total sample of 169 parents. Column 2 excludes parents who read more than 600 minutes (1 from the control group, 1 from the treatment group). Column 3 excludes parents who read more than 450 minutes (1 from the control group, 4 from the treatment group). Column 4 excludes parents who read more than 300 minutes (2 from the control group, 12 from the treatment group). Column 5 excludes parents who did not read at all (14 from the control group, 4 from the treatment group). SD Control Group is the standard deviation of the control group for the respective dependent variable. The effect size is the ratio of the coefficient for the treatment group to the standard deviation of the control group.

Table 1.5: The effect of the treatment on minutes read using covariates, preschool fixed effects, and preschool clustering

	Dependent Variable: Reading Time				
	(1)	(2)	(3)	(4)	(5)
	Baseline: All PACT Parents	Sample with complete information	+Preschool Fixed Effects	+Covariates	+Preschool clustering
Treatment (δ)	88.32***	79.50***	80.09***	83.03***	83.03**
	(18.79)	(20.00)	(19.27)	(21.07)	(26.06)
Constant (α)	63.34***	66.87***	44.98***	-40.3	-40.3
	(9.49)	(10.25)	(14.26)	(89.86)	(103.61)
SD Control Group	87.48	89.89	89.89	89.89	89.89
Effect Size	1.01	0.88	0.89	0.92	0.92
Complete Information	No	Yes	Yes	Yes	Yes
School Fixed Effects	No	No	Yes	Yes	Yes
Covariates	No	No	No	Yes	Yes
Clustering	No	No	No	No	Yes
Observations	169	151	151	151	151

Notes: Robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table shows results for the estimation of equation (1) using "Minutes Read", as outcome, using covariates, preschool fixed effects, and preschool clustering, explained as follows. Column 1 includes all 169 parents. Columns 2-5 exclude parents with missing covariates. The covariates included in column 4 and 5 are all those listed in Table 1.2. SD Control Group is the standard deviation of the control group for the respective dependent variable. The effect size is the ratio of the coefficient for the treatment group to the standard deviation of the control group.

Table 1.6: Survey Response

	Treatment	Control	Total	% Total
All Recruited Parents	84	85	169	100%
Answered Survey 1	80	80	160	95%
Answered Survey 2	59	53	112	66%
Answered Survey 3	56	56	112	66%

Note: The first row shows the number of parents that were recruited to participate in the PACT study, by treatment status. These correspond to all 169 parents that signed the consent forms, participated of the orientation sessions and received the tablet. The number of parents answering surveys 1, 2 and 3 are shown in the last three rows, respectively.

Table 1.7: Percent of Parents in the PACT sample and Parents of Children Entering Head Start in 2010 with various characteristics

	Head Start FACES	PACT
Speaks English at home	73.7	74.4
Education		
Less than HS	31.8	27.5
High school diploma or GED	35.2	17.5
Some college (including 2-year degrees)	25.3	41.9
BA or more schooling	7.8	13.1
Race/ethnicity		
African American	33.0	32.1
Hispanic/Latino	36.0	65.2
Child has disability	14.1	12

The column 'Head Start FACES' presents statistics from national sample of parents of children in Head Start programs on characteristics collected in the Head Start Family and Child Experiences Survey (FACES) on 2010, the latest year for which FACES data is available. The second column exhibits mean characteristics for all the parents participating in the PACT intervention.

Table 1.8: Treatment effect for participants who took each of 3 surveys

Panel 1. Dependent Variable: Reading Time

	(1)	(2)	(3)	(4)
Treatment (δ)	88.32***	83.00***	82.12***	88.88***
	(18.79)	(19.35)	(24.08)	(24.71)
Constant (α)	63.34***	66.51***	83.61***	82.81***
	(9.49)	(9.95)	(14.11)	(13.44)
SD Control Group	87.48	89.02	102.76	100.61
Effect Size	1.01	0.93	0.8	0.88
Sample	All	Survey 1	Survey 2	Survey 3
Observations	169	160	112	112

Panel 2. Dependent Variable: Number of Books

	(1)	(2)	(3)	(4)
Treatment (δ)	16.66***	15.80***	14.29**	15.77**
	(4.50)	(4.69)	(6.08)	(6.10)
Constant (α)	14.78***	15.57***	19.83***	19.59***
	(3.00)	(3.17)	(4.63)	(4.39)
SD Control Group	27.68	28.34	33.72	32.87
Effect Size	0.6	0.56	0.42	0.48
Sample	All	Survey 1	Survey 2	Survey 3
Observations	169	160	112	112

Panel 3. Dependent Variable: Any Reading

	(1)	(2)	(3)	(4)
Treatment (δ)	0.12**	0.09**	0.06	0.04
	(0.05)	(0.04)	(0.05)	(0.04)
Constant (α)	0.84***	0.87***	0.91***	0.93***
	(0.04)	(0.04)	(0.04)	(0.03)
SD Control Group	0.37	0.33	0.3	0.26
Effect Size	0.31	0.26	0.2	0.14
Sample	All	Survey 1	Survey 2	Survey 3
Observations	169	160	112	112

Notes: Robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Column 1: considers all 169 parents (84 in the treatment group, and 85 in the control group). Column 2: considers the 160 parents that answered survey 1 (80 in the treatment group, and 80 in the control group). Column 3: considers the 112 parents that answered survey 2 (59 in the treatment group, and 53 in the control group). Column 4: considers the 112 parents that answered survey 2 (56 in the treatment group, and 56 in the control group). SD Control Group is the standard deviation of the control group for the respective dependent variable. The effect size is the ratio of the coefficient for the treatment group to the standard deviation of the control group.

Table 1.9: Robustness Check: Excluding Outliers

Panel 1. Dependent Variable: Reading Time

	(1)	(2)	(3)	(4)
Treatment (δ)	88.32***	86.48***	72.48***	51.30***
	(18.79)	(15.57)	(13.74)	(10.75)
Constant (α)	63.34***	56.69***	56.69***	52.67***
	(9.49)	(6.85)	(6.85)	(5.62)
SD Control Group	87.48	62.78	62.78	51.16
Effect Size	1.01	1.38	1.15	1
Excluding Parents who read more than	-	600 minutes	450 minutes	300 minutes
Observations	169	167	164	155

Panel 2. Dependent Variable: Number of Books

	(1)	(2)	(3)	(4)
Treatment (δ)	16.66***	17.41***	15.25***	12.23***
	(4.50)	(3.19)	(2.91)	(2.52)
Constant (α)	14.78***	12.17***	12.17***	11.43***
	(3.00)	(1.50)	(1.50)	(1.33)
SD Control Group	27.68	13.77	13.77	12.1
Effect Size	0.60	1.26	1.11	1.01
Excluding Parents who read more than	-	600 minutes	450 minutes	300 minutes
Observations	169	167	164	155

Panel 3. Dependent Variable: Any Reading

	(1)	(2)	(3)	(4)
Treatment (δ)	0.12**	0.12**	0.12**	0.11**
	(0.05)	(0.05)	(0.05)	(0.05)
Constant (α)	0.84***	0.83***	0.83***	0.83***
	(0.04)	(0.04)	(0.04)	(0.04)
SD Control Group	0.37	0.37	0.37	0.38
Effect Size	0.31	0.32	0.31	0.3
Excluding Parents who read more than	-	600 minutes	450 minutes	300 minutes
Observations	169	167	164	155

Notes: Robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Column 1: considers all 169 parents Column 2: excludes parents that read more than 600 minutes (1 from the control group, 1 from the treatment group). Column 3: excludes parents that read more than 450 minutes (1 from the control group, 4 from the treatment group). Column 4: excludes parents that read more than 300 minutes (2 from the control group, 12 from the treatment group). Column 5: excludes parents that did not read at all (14 from the control group, 4 from the treatment group). SD Control Group is the standard deviation of the control group for the respective dependent variable. The effect size is the ratio of the coefficient for the treatment group to the standard deviation of the control group.

Table 1.10: Robustness Check: Covariates, Preschool Fixed Effects and Clustering

Panel 1. Dependent Variable: Reading Time

	(1)	(2)	(3)	(4)	(5)
Treatment (δ)	88.32***	79.50***	80.09***	83.03***	83.03**
	(18.79)	(20.00)	(19.27)	(21.07)	(26.06)
Constant (α)	63.34***	66.87***	44.98***	68.13	68.13
	(9.49)	(10.25)	(14.26)	(101.14)	(131.74)
SD Control Group	87.48	89.89	89.89	89.89	89.89
Effect Size	1.01	0.88	0.89	0.92	0.92
School Fixed Effects	No	No	Yes	Yes	Yes
Covariates	No	No	No	Yes	Yes
Clustering	No	No	No	No	Yes
Observations	169	151	151	151	151

Panel 2. Dependent Variable: Number of Books

	(1)	(2)	(3)	(4)	(5)
Treatment (δ)	16.66***	15.32***	15.11***	16.35***	16.35*
	(4.50)	(4.92)	(4.73)	(5.30)	(8.05)
Constant (α)	14.78***	15.96***	12.05***	16.57	16.57
	(3.00)	(3.28)	(3.23)	(27.13)	(35.25)
SD Control Group	27.68	28.8	28.8	28.8	28.8
Effect Size	0.6	0.53	0.52	0.57	0.57
School Fixed Effects	No	No	Yes	Yes	Yes
Covariates	No	No	No	Yes	Yes
Clustering	No	No	No	No	Yes
Observations	169	151	151	151	151

Panel 3. Dependent Variable: Any Reading

	(1)	(2)	(3)	(4)	(5)
Treatment (δ)	0.12**	0.10**	0.10**	0.12***	0.12**
	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)
Constant (α)	0.84***	0.87***	0.91***	0.69***	0.69***
	(0.04)	(0.04)	(0.05)	(0.18)	(0.18)
SD Control Group	0.37	0.34	0.34	0.34	0.34
Effect Size	0.31	0.3	0.31	0.36	0.36
School Fixed Effects	No	No	Yes	Yes	Yes
Covariates	No	No	No	Yes	Yes
Clustering	No	No	No	No	Yes
Observations	169	151	151	151	151

Notes: Robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Column 1: considers all 169 parents. Columns 2-5: excludes parents with missing covariates. SD Control Group is the standard deviation of the control group for the respective dependent variable. The effect size is the ratio of the coefficient for the treatment group to the standard deviation of the control group.

Table 1.11: Time Preferences Sensitivity

Panel 1. T-test for participation in Survey 2 by treatment status

Variable	Total Sample	Treatment	Control	Difference	*P-value*
Participation in Survey 2 (%)	0.663	0.702	0.624	0.078	0.281

Panel 2. T-test for variation in Survey 2 answers by treatment status

Variable	Survey 2 Sample	Treatment	Control	Difference	*P-value*
Variation in Survey 2 Answers (%)	0.750	0.729	0.774	-0.045	0.542

Panel 3. T-test for More/Less Patient A by treatment status

Variable	Survey 2 Sample	Treatment	Control	Difference	*P-value*
More Patient A (%)	0.500	0.535	0.463	0.071	0.518

Panel 4. T-test for more/less patient B by treatment status

Variable	Survey 2 Sample	Treatment	Control	Difference	*P-value*
More Patient B (%)	0.482	0.492	0.472	0.019	0.836

Notes: 'Participation in Survey 2 (%)' refers to the fraction of parents from the original sample (N=169) who answered Survey 2. 'Variation in Survey 2 Answers (%)' refers to the fraction of parents who did not give the same answer to all questions in the time preferences task.'More Patient A' refers to the fraction of parent whose time discounting (delta) parameter score is above the median. 'More Patient B': refers to the fraction of parents who answered only C or only D in the time preferences task.

Table 1.12: Time Preferences and Treatment Effects. Dependent Variable: Reading Time

	(1) Baseline: All PACT Parents	(2) Took Survey 2
Treatment (δ)	88.32***	82.12***
	(18.79)	(24.08)
Constant (α)	63.34***	83.61***
	(9.49)	(14.11)
SD Control Group	87.48	102.76
Effect Size	1.01	0.8
Observations	169	112

	(3) With Time Preferences A	(4) Less Patient Parents A	(5) More Patient Parents A	(6) With Time Preferences B	(7) Less Patient Parents B	(8) More Patient Parents B
Treatment (δ)	80.33***	139.97***	16.44	82.12***	124.52***	42.26
	(30.30)	(39.73)	(41.81)	(24.08)	(33.25)	(33.95)
Constant (α)	91.78***	75.36***	105.96***	83.61***	68.33***	97.26***
	(17.62)	(14.19)	(30.58)	(14.11)	(12.57)	(24.19)
SD Control Group	112.88	62.03	143.25	102.76	62.95	128.08
Effect Size	0.71	2.26	0.11	0.8	1.98	0.33
Observations	84	42	42	112	54	58

Notes: Robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. With Time Preferences A: excludes parents without variation in the time orientation task answers. Impatient (patient) parents A are those whose score on the time orientation task is below (above) the median. With Time Preferences B: includes parents as patient if they answered always 100% or 66% of the budget sooner. Impatient (patient) parents A are those whose score on the time orientation task is below (above) the median. SD Control Group is the standard deviation of the control group for the respective dependent variable. The effect size is the ratio of the coefficient for the treatment group to the standard deviation of the control group.

Table 1.13: Number of inconsistent choices by treatment status

Number of Inconsistent choices	Group: Treatment	Group: Control	Total
0	34	27	61
1	8	9	17
2	7	6	13
3	4	4	8
4	3	4	7
5	3	3	6
Total	59	53	112

Notes: A choice is classified as consistent if it allocates the same or a larger budget share to the future payment date as a previous choice with a lower gross rate. As the gross rate increases, the law of demand dictates a weakly decreasing allocation to the earlier time point.

Table 1.14: Time Preferences and Treatment Effects for Consistent Parents. Dependent Variable: Reading Time

	(1) Baseline: All PACT Parents & Consistent	(2) Took Survey 2 & Consistent
Treatment (δ)	88.32***	81.74***
	(18.79)	(26.42)
Constant (α)	63.34***	90.17***
	(9.49)	(15.93)
SD Control Group	87.48	108.14
Effect Size	1.01	0.76
Observations	169	99

	(3) With Time Preferences A & Consistent	(4) Less Patient Parents A & Consistent	(5) More Patient Parents A & Consistent	(6) With Time Preferences B & Consistent	(7) Less Patient Parents B & Consistent	(8) More Patient Parents B & Consistent
Treatment (δ)	79.66**	136.03***	21.55	81.74***	121.33***	46.87
	(34.44)	(45.32)	(46.39)	(26.42)	(36.86)	(36.35)
Constant (α)	102.33***	85.13***	114.37***	90.17***	73.41***	103.06***
	(20.66)	(17.76)	(33.09)	(15.93)	(15.04)	(25.70)
SD Control Group	120.56	66.89	147.68	108.14	67.48	131.11
Effect Size	0.66	2.03	0.15	0.76	1.8	0.36
Observations	71	34	37	99	46	53

Notes: Robust standard errors in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We classified a parent to be systematically consistent if the number of inconsistent choices is less or equal to three. This is arbitrary, but the results hold for any other cutoff. With Time Preferences A: excludes parents without variation in the time orientation task answers. Impatient (patient) parents A are those whose score on the time orientation task is below (above) the median. With Time Preferences B: includes parents as patient if they answered always 100% or 66% of the budget sooner. Impatient (patient) parents A are those whose score on the time orientation task is below (above) the median. SD Control Group is the standard deviation of the control group for the respective dependent variable. The effect size is the ratio of the coefficient for the treatment group to the standard deviation of the control group.

CHAPTER 2

THE VANISHING BENEFITS OF DELAYED PRIMARY SCHOOL ENROLLMENT

(With P. Celhay)

2.1 Introduction

Nearly all education systems have cutoff dates for school eligibility. These cutoff dates are important because some students are older than others when they begin school. This has been used by researchers to study the effects of school starting age. The literature has found consistently that older students outperform younger students on in-school tests (by roughly 0.4σ or higher for fourth grade or below, and 0.2σ for eighth grade, according to Bedard and Dhuey (2006); Puhani and Weber (2007); Datar (2006); McEwan and Shapiro (2008); Elder and Lubotsky (2009); Smith (2009). If this effect[1] is significant in early primary grades but then dissipates over time, this phenomenon, while interesting, may not be particularly important. On the other hand, if the effects propagate themselves through the human capital accumulation process into later life, they may have important implications for adult outcomes and productivity.

Due to a lack of quality data, there is less evidence on longer-term outcomes like high school completion, transitions to post-secondary education, and earnings, which are arguably more relevant in shaping students' future life trajectories than test scores in elementary school. Some exceptions are Dobkin and Ferreira (2010) for the US (California and Texas), Black et al. (2011) for Norway and Kawaguchi (2011) for Japan.

This paper examines whether initial benefits of delayed school enrollment persist in the

1. The age at test effect and schooling effect as defined in Black et al. (2011) and McEwan and Shapiro (2008) are indistinguishable from each other when analyzing in-school tests. We work with this restriction for our in-school measures, and interpret the effects of delayed enrollment as the sum of these two effects as in Smith (2009).

long run. Like previous articles, we employ a regression discontinuity framework, exploiting exogenous variation in age in first grade created by enrollment cutoff dates. Unlike prior research, we carefully track students year by year as they progress through primary and secondary school, finish high school, and transition to post-secondary education, based on administrative student-level data from Chile. We are thus able contribute to this literature by assessing precisely when the impact of delayed enrollment arises and its related consequences. The rich nature of the data also allows us to examine not only student achievement scores and GPA, but also other non-test-score outcomes such as attendance, on-time grade progression, school switching, pursuing a vocational track, schooling attainment, and post-secondary enrollment.

Our findings indicate that there are initial positive effects of a one-year delay on GPA and test scores, which decrease monotonically as students grow older. In fact, those effects vanish by the end high school, have little (if any) effect on national college admission test scores, have no effect on the probability of enrollment in any higher education institution, let alone college or in flagship selective universities. Overall, our results suggest that school-starting age does not have significant effects on later life outcomes.

The remainder of the paper is organized as follows. Section 2.2 outlines our empirical strategy, while Section 3.4 describes the data we use in our analysis. Section 2.4 presents and discusses the results. Section 2.5 concludes.

2.2 Empirical Strategy

The estimation of a naive equation that relates student' outcomes to school starting age could in principle provide us 'age effect' parameters. However, unobserved factors that affect outcomes may also drive the motivation to delay entry. Thus OLS will not identify age effects. For example, parents who value education more might enroll their children at a younger age. In that case, OLS would deliver downward biased estimates of school starting age. It could also be the case that less educated parents send their children to school earlier since it provides free childcare. In that scenario the OLS estimate would be upward biased.

The existing cutoff dates for school eligibility in Chile generate exogenous variation in the school starting age, which allows for the estimation of age effects on outcomes of interest. Chile's academic year runs from March to December. The official enrollment cutoff for first grade is April 1, but in practice, the Ministry of Education allows schools to implement cutoff dates as late as July 1st. That late cutoff date is actually the most commonly used by Chilean schools. This implies that a prospective student who is not 6 years old by June 30th is supposed to delay first grade enrollment by one year.

Consider Figure 2.13 for illustrative purposes. The horizontal axis represents a time-line for birthdays over years T and T+1. In the middle of both years, the dashed line marks the cutoff date of July 1. The graph shows how the delayed enrollment works for three adjacent cohorts. Children in the vicinity defined by 'A,' were born only days apart, but those falling on the right side delayed enrollment by one year because they just missed the cutoff date. Therefore, those born before July 1 in year T are supposed to enroll in first grade in year T+6, i.e. when 6 years old. Those born on or after July 1 in year T delay their entrance to school and enroll with the next cohort, in year T+7, when they are 7 years old. Similarly, children born before July 1 in year T+1 (and after July 1 in year T), enroll in first grade in year in T+7 (6 years old), while those born after July 1 in year T+1 enroll in first grade in T+8, at 7 years old. As such, these cutoffs allocate children into different, adjacent, entry years and therefore, students have very similar birth dates but are markedly different in the age when they start first grade.[2]

To estimate effects of delayed enrollment, we employ a regression discontinuity (RD) design, where the running variable is the birth date of student i denoted by B_i. We perform our analysis in the spirit of Figure Y but considering four possible cutoff dates: the first day of April, May, June, and July. Given that there may be imperfect compliance with the cutoff

2. In addition, the design implies that those falling in the left side of the cutoff are the youngest in their grade, while those on the right side are the oldest.

dates, and hence the probability of 'treatment' may jump by less than one at the thresholds, we estimate a fuzzy RD relying on the following equations

$$Y_i = \beta_0 + \beta_1 A_i + \beta_2 X_i + f(B_i) + \varepsilon_i \tag{1}$$

$$A_i = \alpha_0 + \sum_{j=1}^{4} \alpha_j D_{ij} + \alpha_5 X_i + f(B_i) + v_i \tag{2}$$

where i indexes students and j indexes the four cutoff dates given by first day in April (1), May (2), June (3), and July (4). D_{ij} is an indicator for the student i's birth date (B_i) being equal to or exceeding the cutoff j. Y_i represents a particular student outcome such GPA, attendance rate, persistence, tests scores, among others. Xi includes a set of predetermined variables: gender, socioeconomic status, parental education, income, class size, rural versus urban school, type of school (public, voucher, private), school location, etc. In practice, these control variables have very little effect on our discontinuity estimates and serve mainly to increase precision. We describe our vector of outcomes and covariates in detail in section 3.4.

The function f(B) captures smooth, seasonal effects of birth dates on student outcomes. It is a smooth function of the birth date, which is the forcing variable determining delayed enrollment. The central assumption underlying the RD design is that we have correctly specified the function, f(B). We consider both parametric and non-parametric functions of birth date and explore the robustness of our findings to a variety of functional form assumptions.

We first specify a flexible parametric model by including higher order polynomials of Bi, which are allowed to vary on either side of the discontinuities. We consider specifications that use linear, quadratic, and cubic trends in the date of birth, as well as linear, quadratic and cubic splines where the trends are fully interacted with D_{ij}, as presented below.

$$f(B_i) = \sum_{k=1}^{3} \delta^k {B_i}^k + \sum_{j=1}^{4} \sum_{k=1}^{3} \delta_j^k D_{ij} (B_i - \overline{B}_{ij})^k \tag{3}$$

For our non-parametric specifications we follow Hahn et al (2001) and Porter (2003), and estimate a local (kernel-weighted) linear regression to estimate the left and right limits of the discontinuity. Given that we have different cutoffs, we proceed by estimating the local regressions using different sets of data points near the respective thresholds. The different intention-to-treat parameters are computed as the difference between the left and right limits of these regressions at each of the birth date cutoffs. We basically estimate the following local linear regression to the left and right of our four cutoffs, where the difference between the two (at the respective cutoffs) is the estimated intention-to-treat effects;

$$\sum_{i \in \{B_i \geq \overline{B}_{ij}\}}^{N_j} K\left(\frac{B_i - \overline{B}_{ij}}{h_j}\right) (Y_i - \alpha_{w,x} - \beta_{w,x} B_i)^2 \tag{4}$$

We estimate this in one step using a rectangular kernels.[3] To get the different LATEs, we divide the ITTs by the respective difference in treatment receipt rates for both groups within a close neighborhood around the cutoff, defined by a bandwidth h_j (which could in principle differ for the various cutoffs). In general, choosing a bandwidth in a nonparametric estimation involves finding an optimal balance between precision and bias. As the number of available observations increases, it becomes possible to use an increasingly small bandwidth since linear regressions can be estimated relatively precisely over even a small range of values of B_i.

Having data on the entire population of students allows us to pick small bandwidths and still have a high number of observations to perform the RD analysis. This makes the bias-variance trade-off negligible in our setup. The local nonparametric strategy substantially

3. Our results are not sensitive to the choice of kernel.

reduces the chances of bias and it has negligible costs in terms of precision given the statistical power we have near the thresholds. This makes the nonparametric approach preferred to the parametric estimation procedure.

We follow Ludwig and Miller (2007) and examine our results for a range of candidate bandwidths. Additionally, we conduct standard RD robustness checks. We test for (i) manipulation of birth timing by checking the frequency density along the lines of McCrary (2008), and (ii) examine whether pre- treatment background characteristics vary continuously around the thresholds.

As we follow students over time, our sample is naturally subject to suffer from attrition. If the number of observations drops discontinuously at the thresholds, that may change the composition of the remaining observations and impose a threat to the internal validity of our results. In other words, this differential attrition may introduce differences in ability among students born just before and just after a cutoff, which could bias our RD estimates.[4]

To overcome the problem of selective attrition, we present the primary and high school completion rates for our initial population of students. These results are free of attrition because we track the original population of 1st graders that started school at the same time and compute for all of them a variable indicating whether they completed each level on time, late, or dropped out.

Finally, the average effect of delayed enrollment near the birth cutoffs may hide important distributional effects. For example, the potential advantages of being older when starting school may be larger for children from disadvantaged socioeconomic backgrounds, whose outcomes tend to lie on the lower end of the distribution. We explore the existence of those effects by using the methodology developed by Frandsen, Frolich, and Melly (2012) to compute quantile treatment effects. Following their procedure, we use local distribu-

4. Following Lee and Lemieux (2010), we first test for selective attrition linked to treatment status by examining discontinuities in the density, conditional on survival. Our census data allows us to track students that did not advance grades timely, and therefore we observe their outcomes even if they repeat (even twice). The Online Appendix presents the levels of attrition, the results without imputations, the results with imputations, results performing median regressions, and results considering bounds on the outcomes.

tion regression to estimate the marginal distributions of the potential outcomes. We then compute the local quantile treatment effect as the difference between the inferred marginal distributions of the potential outcomes for compliers at a particular quantile (or in other words, as the difference between different 'local Wald ratios'). This method allows for direct investigation of the effect of delaying enrollment on the lower end of the distribution of our different outcomes.

2.3 Data

We use different information sources to implement our empirical strategy. Our primary data source is the Chilean Students' Registry, which is an administrative dataset maintained by the Ministry of Education that covers the complete population of students since 2002. Using this information, we study two sets of student cohorts.

The first set is students enrolled for first-time in first grade in 2002, 2003, and 2004, which we follow until they finish primary school in 8th grade. We augment these data with the information from the SIMCE test administered yearly to all Chilean fourth since 2005. This allows us to supplement the data with tests scores for language and mathematics and with additional information collected from the parental surveys associated with the SIMCE test. [5] administered to all Chilean fourth graders every year since 2005. This allows us to supplement the data with tests scores for language and mathematics and with additional information collected from the parental surveys associated to the SIMCE test.

To measure results as students transition from primary to secondary school, we analyze a second set of cohorts, who were 1st graders in 1996, 1997, and 1998. We observe them when they were 8th graders in 2003, 2004, and 2005, and follow them until they finish high school. We supplement this data with two additional sources: (i) information coming from

5. SIMCE stands for National System of Education Quality Measurement in Spanish. The SIMCE test consists in a battery of tests on the subjects of math and language administered to all Chilean fourth graders. The SIMCE test is also applied with certain regularity in 8th grade and 10th grade. For more on the SIMCE test and a calendar of its application, the interested reader can visit www.simce.cl.

the Chilean National Admission Test[6] which high school graduates take at the end of their senior year, and (ii) administrative information on higher education enrollment that allows us to determine the probability of enrolling in any higher education institution, college, and in a flagship selective universities.

Table 2.1 shows descriptive statistics for the three cohorts of first graders in 2002, 2003 and 2004 respectively. More than 250,000 children were enrolled each of these years. All three cohorts share common characteristics as presented in panel A.[7]

The average school starting age is 6.3 years old. Students' parents have 11.5 years of schooling on average. About half (49%) of the children are female, and about 90% of the students attend a school located in the same comuna of their residence. The average first grade class size is about 32 students. Approximately 40% of the students attend schools located in the Metropolitan Region (the national capital), and more than 13% of them are in rural areas. In addition, almost half of the students go to public schools, more than 40% go to voucher schools and less than 8% of them attend private schools. All these predetermined variables are smoothly distributed across the thresholds as presented and described in Section 2.4.

Panel B of Table 1 exhibits a set of academic outcomes for the first graders. It shows that that more than 95% of them continue 2nd grade; have an average annual attendance rate of approximately 92%; and a GPA of 61 on a 10-70 scale, where 40 is needed to pass. Table 2 presents descriptive statistics for the students' outcomes as they progress through primary school. Persistence and attendance rates, and the standardized GPAs are computed considering all students who were part of the respective grade. Persistence rates are fairly stable across cohorts and years. Attendance rates fluctuate a little, with a tendency to

6. This test, known as the PSU (Prueba de Selección Universitaria) is a national exam taken at the end of senior year. This test is mandatory for admission to the most important universities and also required by many private universities and professional institutes for admission. For many Chilean universities, the PSU score combined with GPA are the sole determinants of admission.

7. We work with a raw data set in which we drop observations with invalid identification codes and observations for which we do not have the date of birth. The percentage of observations we lose is around 2%-3%.

increase towards the middle of primary school and then decline towards the ends. GPAs, standardized within school for each grade, increase over time as grade repeaters are left behind year by year (note that their grades are used to compute the GPA within schools). The standardized test scores taken as part of the SIMCE evaluation are also presented in the table. Finally, we include dummies indicating whether students finished 8th grade on time (8 years), late (within 10 years), or dropped out (which can not be disentangling from those who completed 8th grade in more than 10 years, but we expect that to be a very small fraction of the total). About 75% of the students who started school in 2002, 2003, and 2004 finished on time, 17% finished late, and 8% dropped out.

Figure 2.1 is the empirical representation of Figure X. It graphs first grade enrollment age against birth date for our first set of cohorts. The dots show age means within birthday cells and reveal sharp increases in enrollment age on the first days of April, May, June, and July. As indicated before, the July 1 threshold is revealed to be the most common one. The solid line plots fit values from local linear regressions for first grade enrollment age on both sides of the birth date cutoff.

2.4 Results

Our findings indicate that there are initial, important, positive effects of a one-year delay on GPA. In first grade the intention-to-treat (ITT) estimate is about 0.3 sd. In 4th grade, the intention-to-treat estimate is about 0.2 sd, a result that is also seen in national standardized test scores, also included in Figure 5. Results decrease monotonically as students age. In fact, those effects vanish by the end high school, as shown in Figure 7. Figures 8 and 9 show results for attendance rates. The intention-to-treat estimate appears again to be important in 1st and, though decreasing, in 2nd grade. However, this effect disappears again and this time more rapidly; in 3rd grade there are not any effects, and they don't appear again later on (there seems to be some cohort effects though but of small magnitude).

The effects on persistence are exhibited in Figures 10 and 11. The ITTs follow a pattern that resembles that of GPA. There are effects in early grades that tend to stabilize in older grades and vanish in high school. Once they start high school, Chilean students can choose between pursuing vocational and academic education. The vocational track consists of a curricular program designed to prepare students for work, whereas academic education is designed to prepare students for tertiary education. Lower socioeconomic status students tend to pursue the vocational track as documented by Farias (2013) among others. When analyzing differences in pursuing a vocational track in high school, as illustrated by Figure 12, the previously described tendency persists; there are no differences near the birth date cutoffs.

Figure 13 presents the differences in schooling attainment. The first two graphs exhibit the probability of completing primary school on time (within 8 years) and with a delay of up to 2 years (e.g. within 10 years). The first thing to note is that there is not any differences in primary school attainment by birth date, which can be seen by the offsetting nature of both graphs. The fraction of children not finishing or taking more than 10 years to do so is so tiny that it is not presented here. The second takeaway is that delayed enrollment matters for timely primary school completion. Students to the right of the cutoff (i.e. those who delayed enrollment) have a higher likelihood of finishing on time (of about 5 percentage points, the ITT). A similar phenomenon is depicted in the three next graphs for high school completion, but again, magnitudes remain much smaller.

Finally, outcomes related to the transition to postsecondary education are included in Figure 14. The graphs suggest that there is an effect of delayed enrollment on taking the PSU test on time. This effect is offset by students taking the PSU late (up to two years later), which is consistent with the result found for high school completion described above. Similarly, there are effects in PSU scores (math) if taken on time, but again this is offset when we consider those who take it late. The probability of higher education enrollment does not vary significantly near the thresholds, which suggests that there are no effects on the probability of enrolling in a selective university or in flagship elite institutions.

Quantile Treatment Effects. We estimated the local treatment effects of delaying enrollment on grades in first grade, on test scores in fourth grade, and on the likelihoods of completing primary school, high school, taking the PSU test, the probability of enrolling in any higher education institution, probability of college enrollment, and probability of enrolling in flagship selective universities. The estimated local quantile treatment effects are plotted in Figure Z. The figure shows the point estimates for each quantile index, as well as pointwise (in the quantile index) 90% confidence intervals based on the asymptotic distribution. We find that the effects of delayed enrollment are higher for the lower end of the distribution of grades and test scores, but that it tends to become constant across quantiles as students age.

2.4.1 Robustness Checks

The main identifying assumption behind our strategy is that there are no other changes occurring at the birth date thresholds that could confound our analysis. We test for these differences by regressing a host of different covariates on the different cutoff dates for school eligibility and a smooth function of our running variable birthday. The results of these regressions are presented in Table X. The reported coefficients are for the variable 'Delayed,' defined as 1 for children who delayed enrollment. The coefficients indicate that gender, parental education, household income, class size, residence location (comuna), school vulnerability index, school location (comuna), rural versus urban area, and school type all evolve smoothly through the birth date threshold. This holds for different local-linear bandwidths and polynomial functions of day of birth.

Figure Z provides a visual check of these results by showing how the same set of covariates behave when plotted against birth date. The dots plot the mean of each variable within birthday bins, and the solid line plots fitted values from local linear regressions on both sides of the birth date cutoff. The graphs confirm that different observed characteristics display a smooth behavior along the time-line, which suggests that unobserved characteristics may display a similar behavior.

We further test for differences in unobserved characteristics by examining whether there is manipulation of birth dates near the cutoffs. For example, it could be the case that more motivated parents planned the timing of their children's birth in order for them to be older when enrolling in primary school. We employ McCrary's (2008) density test to examine the existence of sorting in the vicinity of birth date threshold. The test indicates that X gives the discontinuity estimate with associated standard errors of Y (and hence non-significant). Figure W presents a graphical illustration of this test, which shows that the density varies continuously over birth-dates with no significant discontinuity around the threshold.

Finally, we verified that our estimates are generally robust when using a broader set of bandwidth choices. Online Appendix Tables A and B present local linear regressions with bandwidths of 30, 15, 7, and even 1 day below and above the birthdate cutoffs.

2.5 Conclusions

This paper examines whether initial benefits of delayed school enrollment persist in the long run. We exploit exogenous variation in the first grade enrollment age created by Chile's enrollment cutoff dates and follow students until they finish high school, take the national admission test and transition to post-secondary education. All these outcomes are more relevant in shaping the students' future life trajectories than are scores in elementary school.

The rich nature of our data allows us to examine not only student achievement scores and GPA, but also other outcomes such as attendance, persistence, pursuing a vocational track, attainment, and postsecondary enrollment. Regression-discontinuity estimates, based on birth dates cutoffs, show that there are initial positive effects of a one-year delay on GPA, test scores, attendance, and persistence, that decrease monotonically as students grow older. Those effects vanish by the end high school, have little (if any) effect on the PSU test, have no effect on the probability of enrollment in any higher education institution, probability of college enrollment, and probability of enrollment in flagship selective universities. Overall, our results suggest that school starting age does not have significant effects on the outcomes of young adults.

These findings indicate that delaying the school starting age does not have significant effects on the outcomes of young adults; and therefore, provide evidence against the counterintuitive conclusion that children should start school later.

2.6 References

Bedard, K. and E. Dhuey (2006, November). The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects. The Quarterly Journal of Economics 121(4), 14371472.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2011, May). Too Young to Leave the Nest? The Effects of School Starting Age. The Review of Economics and Statistics 93(2), 455467.

Datar, A. (2006, February). Does delaying kindergarten entrance give children a head start? Economics of Education Review 25(1), 4362.

Dobkin, C. and F. Ferreira (2010, February). Do school entry laws affect educational attainment and labor market outcomes? Economics of Education Review 29(1), 4054.

Elder, T. E. and D. H. Lubotsky (2009). Kindergarten Entrance Age and Childrens Achievement: Impacts of State Policies, Family Background, and Peers. Journal of Human Resources 44(3).

Hahn, J., P. Todd, and W. Van der Klaauw (2001, January). Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. Econometrica 69(1), 20109.

Kawaguchi, D. (2011, June). Actual age at school entry, educational outcomes, and earnings. Journal of the Japanese and International Economies 25(2), 6480.

Lee, D. S. and T. Lemieux (2010, June). Regression Discontinuity Designs in Economics. Journal of Economic Literature 48(2), 281355.

Ludwig, J. and D. L. Miller (2007, 02). Does Head Start Improve Childrens Life Chances? Evidence from a Regression Discontinuity Design. The Quarterly Journal of Economics 122(1), 159208.

McCrary, J. (2008, February). Manipulation of the running variable in the regression discontinuity design: A density test. Journal of Econometrics 142(2), 698714.

McEwan, P. J. and J. S. Shapiro (2008). The Benefits of Delayed Primary School Enrollment: Discontinuity Estimates Using Exact Birth Dates. Journal of Human Resources 43(1).

Puhani, P. A. and A. M. Weber (2007, August). Persistence of the School Entry Age Effect in a System of Flexible Tracking. IZA Discussion Papers 2965, Institute for the Study of Labor (IZA).

Smith, J. (2009, November). Can Regression Discontinuity Help Answer an Age-Old Question in Education? The Effect of Age on Elementary and Secondary School Achievement. The B.E. Journal of Economic Analysis & Policy 9(1), 130.

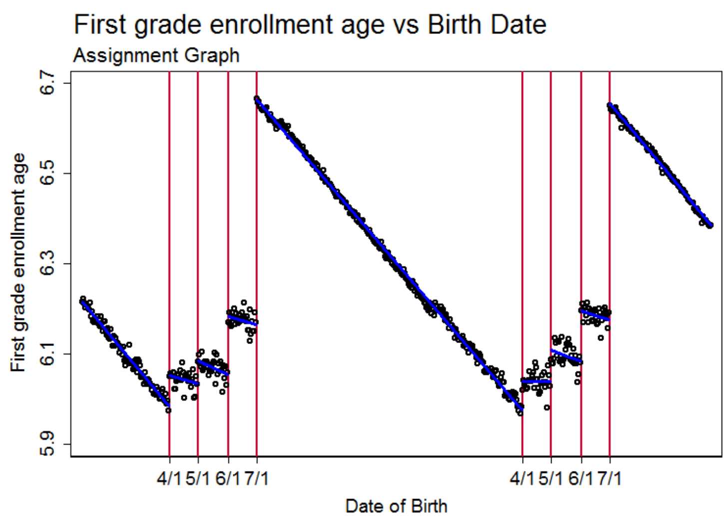
2.7 Figures and Tables

Table 2.1: First graders 2002-2003-2004

	Cohort		
	1st Grade 2002	1st Grade 2003	1st Grade 2004
Observations	259,155	254,763	253,647
Age	6.275	6.282	6.292
Female (%)	0.489	0.488	0.488
Vulnerability Index	27.981	27.870	27.173
Class size	32.039	31.743	31.879
School in Capital Region (%)	0.383	0.388	0.390
Rural School (%)	0.136	0.134	0.131
Public School (%)	0.512	0.492	0.482
Voucher School (%)	0.416	0.429	0.435
Private School (%)	0.072	0.079	0.077
Same Comuna School-Residence (%)	0.908	0.906	0.866
Outcomes (1st Grade)			
Persist 1st grade (%)	0.973	0.952	0.949
Repeat 1 ist grade (%)	0.015	0.035	0.038
Retired 1st grade (%)	0.012	0.013	0.013
Attendance 1st grade	92.793	92.681	92.450
GPA 1st grade	61.330	61.033	60.897

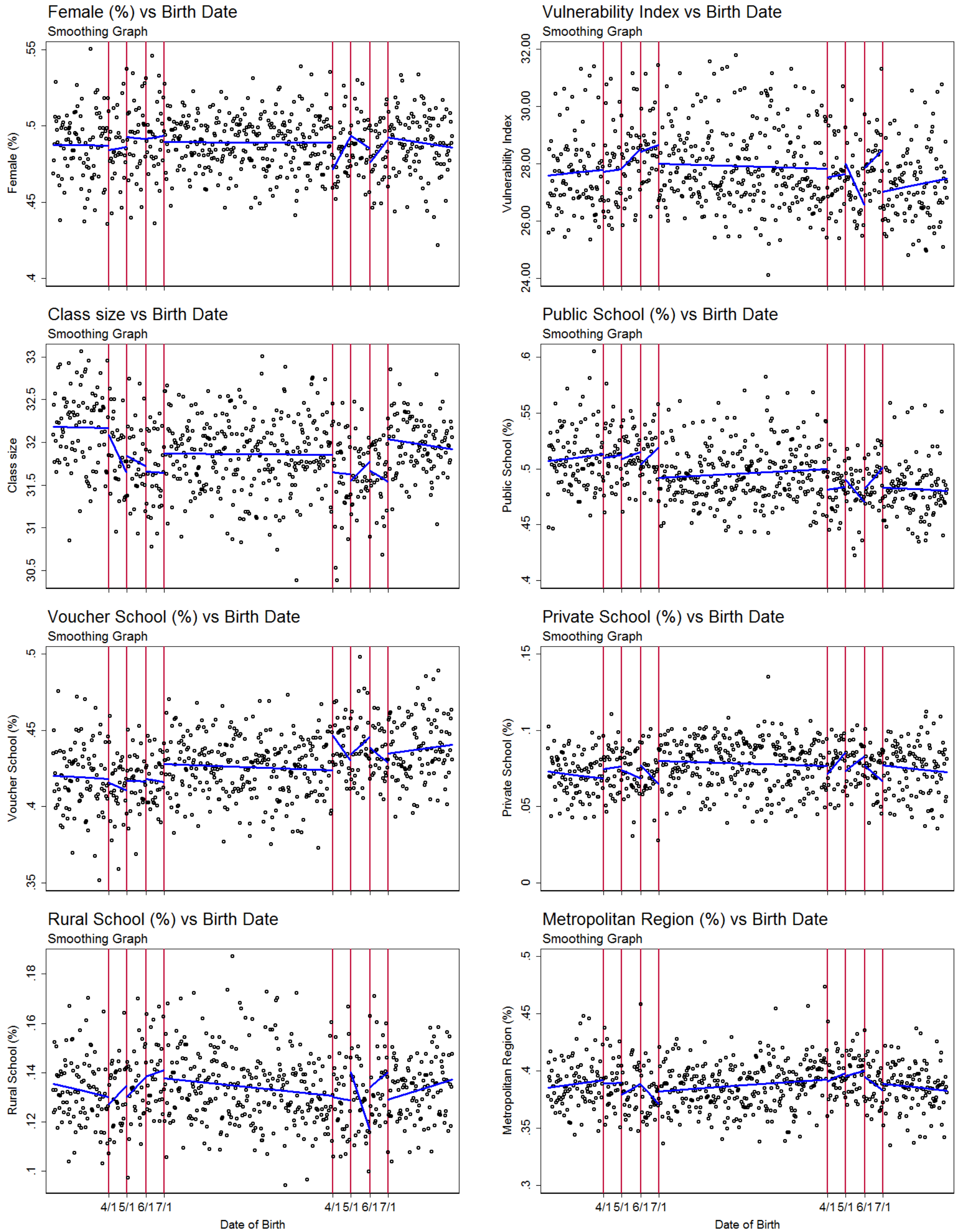
Notes: Table 2.1 shows descriptive statistics for the three cohorts of first graders in 2002, 2003 and 2004 respectively. More than 250,000 children were enrolled each of these years. We work with a raw data set in which we drop observations with invalid identification codes and observations for which we do not have the date of birth. The percentage of observations we lose is around 2%-3%.

Figure 2.1: 1st graders 2002-2003-2004: Enrollment Age



Notes: Figure 2.1 graphs first grade enrollment age against birth date for our first set of cohorts. The dots show age means within birthday cells and reveal sharp increases in enrollment age on the first days of April, May, June, and July. As indicated before, the July 1 threshold is revealed to be the most common one. The solid line plots fit values from local linear regressions for first grade enrollment age on both sides of the birth date cutoff.

Figure 2.2: First Graders: Covariates Near Thresholds



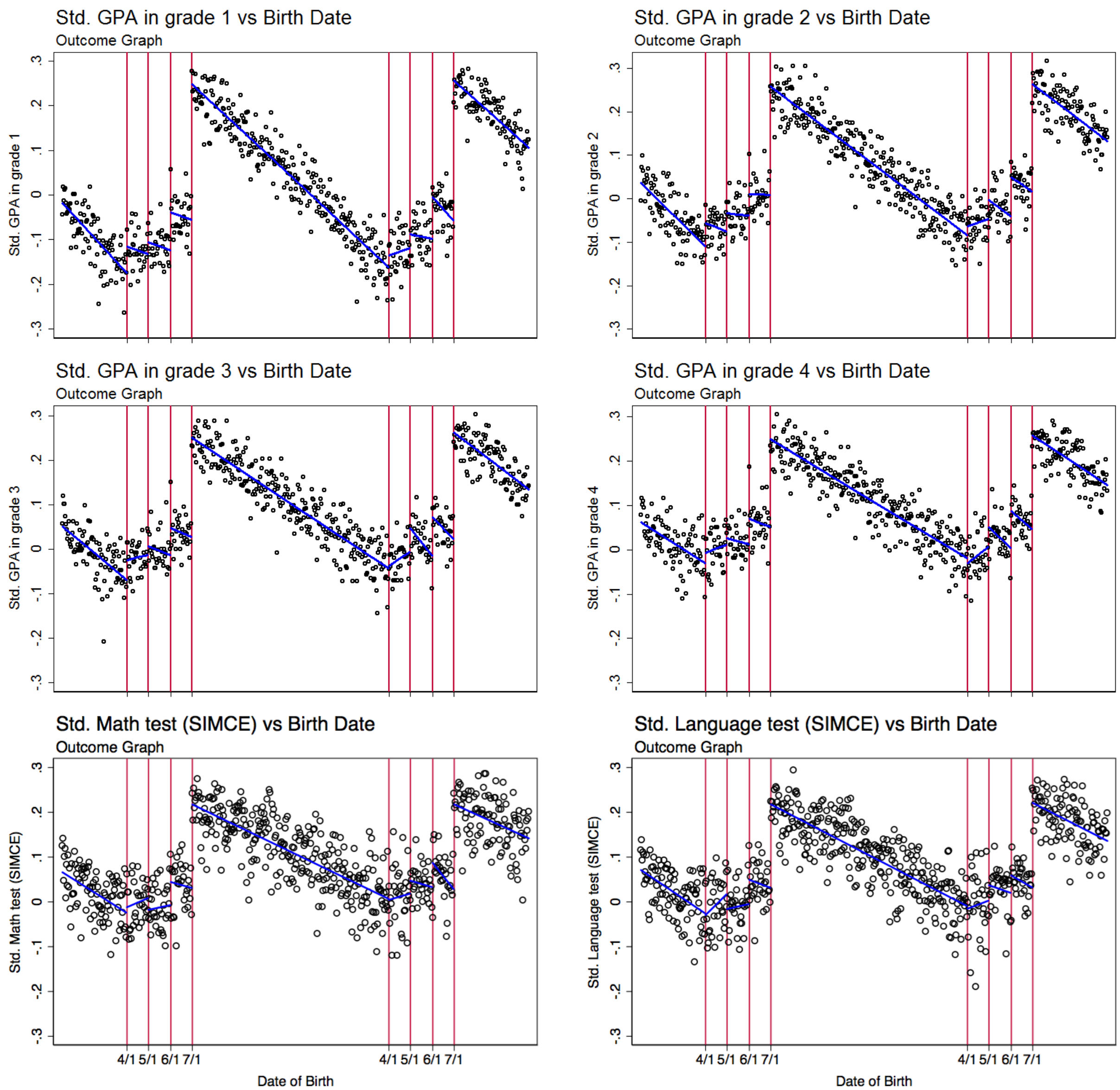
Notes: This Figures show how covariates behave when plotted against birth date. The dots plot the mean of each variable within birthday bins, and the solid line plots fitted values from local linear regressions on both sides of the birth date cutoff. The graphs confirm that different observed characteristics display a smooth behavior along the time-line.

Table 2.2: Following First graders 2002-2003-2004 until 8th grade

Outcomes	Cohort 1st Grade 2002	1st Grade 2003	1st Grade 2004
Persistence			
1st grade	0.973	0.952	0.949
2nd grade	0.953	0.967	0.966
3rd grade	0.971	0.974	0.970
4th grade	0.976	0.976	0.977
5th grade	0.966	0.966	0.965
6th grade	0.965	0.965	0.968
7th grade	0.954	0.964	0.956
8th grade	0.976	0.975	0.968
Attendance			
1st grade	92.793	92.681	92.450
2nd grade	93.754	93.939	93.600
3rd grade	94.242	94.008	93.909
4th grade	94.224	94.166	94.091
5th grade	94.091	94.040	93.463
6th grade	94.159	93.649	93.305
7th grade	93.415	93.180	93.575
8th grade	93.092	93.489	92.054
Standardized GPA			
1st grade	0.008	0.010	0.020
2nd grade	0.051	0.062	0.061
3rd grade	0.071	0.081	0.082
4th grade	0.087	0.094	0.097
5th grade	0.119	0.125	0.127
6th grade	0.139	0.145	0.150
7th grade	0.151	0.160	0.154
8th grade	0.142	0.149	0.139
SIMCE (4th Grade)			
Math test score	0.080	0.095	0.096
Language test score	0.081	0.087	0.088
Mother's schooling	11.200	11.462	11.543
Father's Schooling	11.279	11.571	11.618
Household income	0.793	0.759	0.741
Attainment (8th Grade)			
On time	0.753	0.757	0.742
Late	0.164	0.161	0.176
Drops out	0.083	0.082	0.081

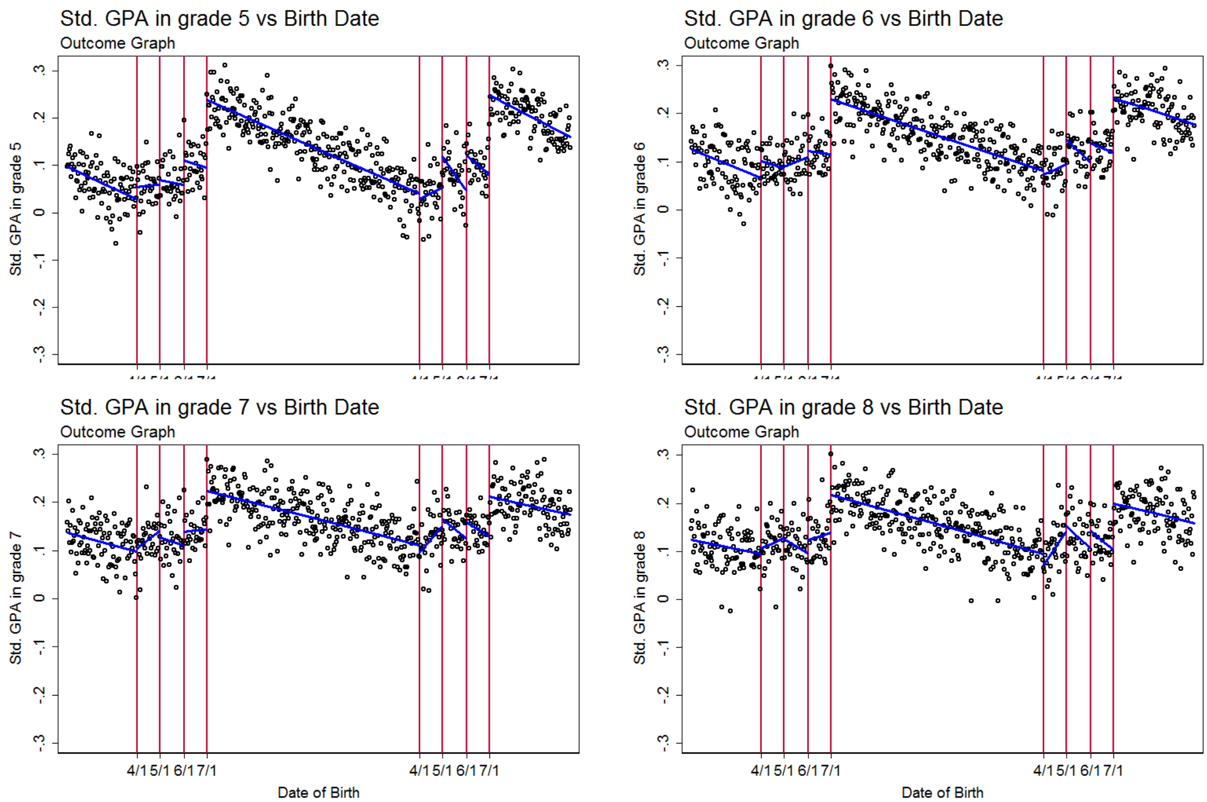
Notes: This Table presents descriptive statistics for the students' outcomes as they progress through primary school. Persistence and attendance rates, and the standardized GPAs are computed considering all students who were part of the respective grade.The standardized test scores taken as part of the SIMCE evaluation are also presented in the table, as well as dummies indicating whether students finished 8th grade on time (8 years), late (within 10 years), or dropped out.

Figure 2.3: Outcomes during Primary School (1th to 4th grade): Std. GPA & Std. Test Scores



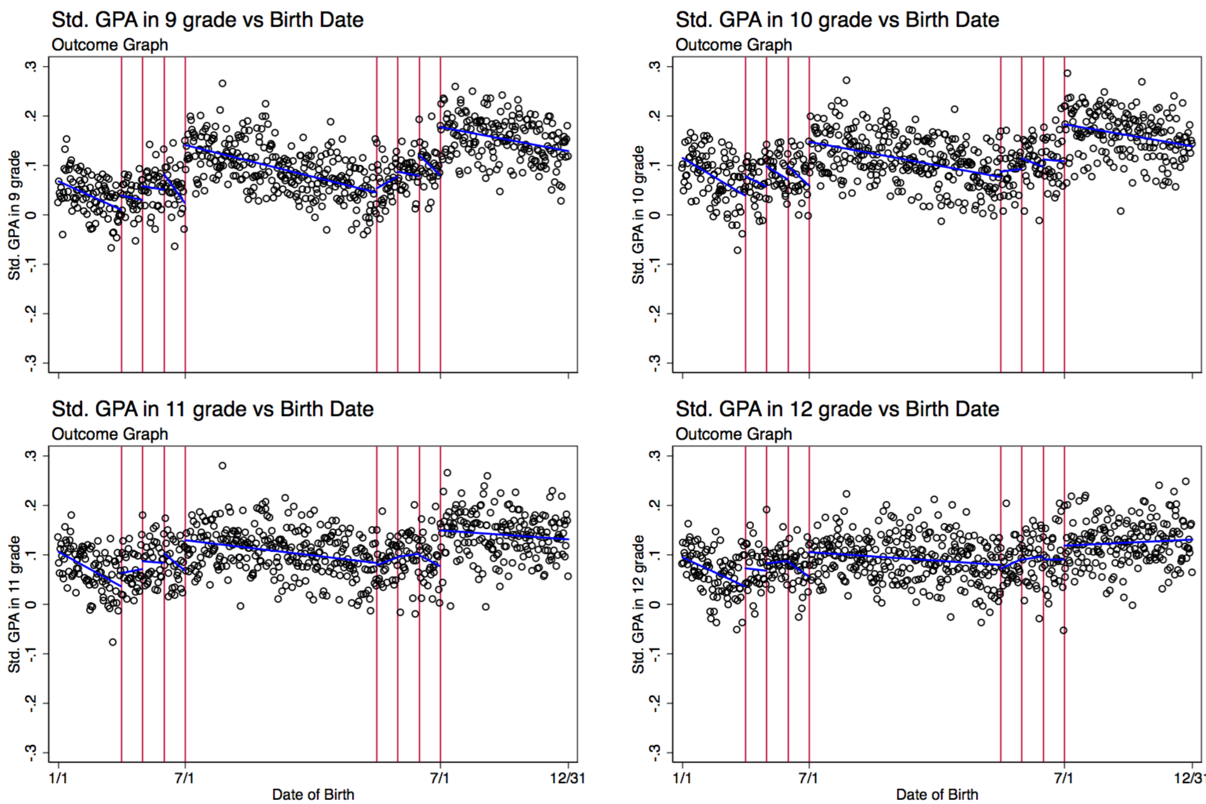
Notes: This Figures show how the respective outcomes behave when plotted against birth date. The dots plot the mean of each variable within birthday bins, and the solid line plots fitted values from local linear regressions on both sides of the birth date cutoff.

Figure 2.4: Outcomes during Primary School (5th to 8th grade): Std. GPA



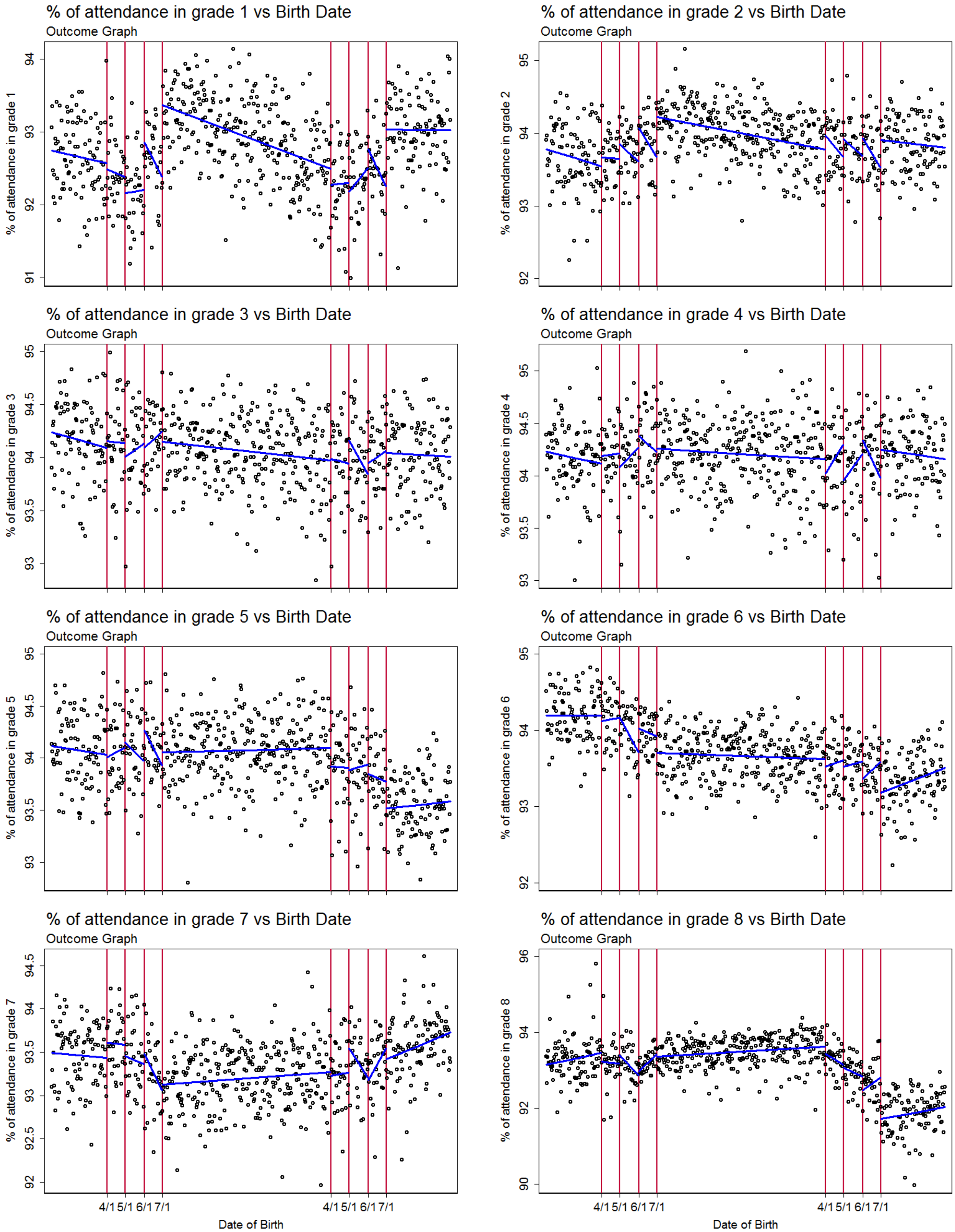
Notes: This Figures show how the respective outcomes behave when plotted against birth date. The dots plot the mean of each variable within birthday bins, and the solid line plots fitted values from local linear regressions on both sides of the birth date cutoff.

Figure 2.5: Outcomes during High School (until 12th grade): Std. GPA



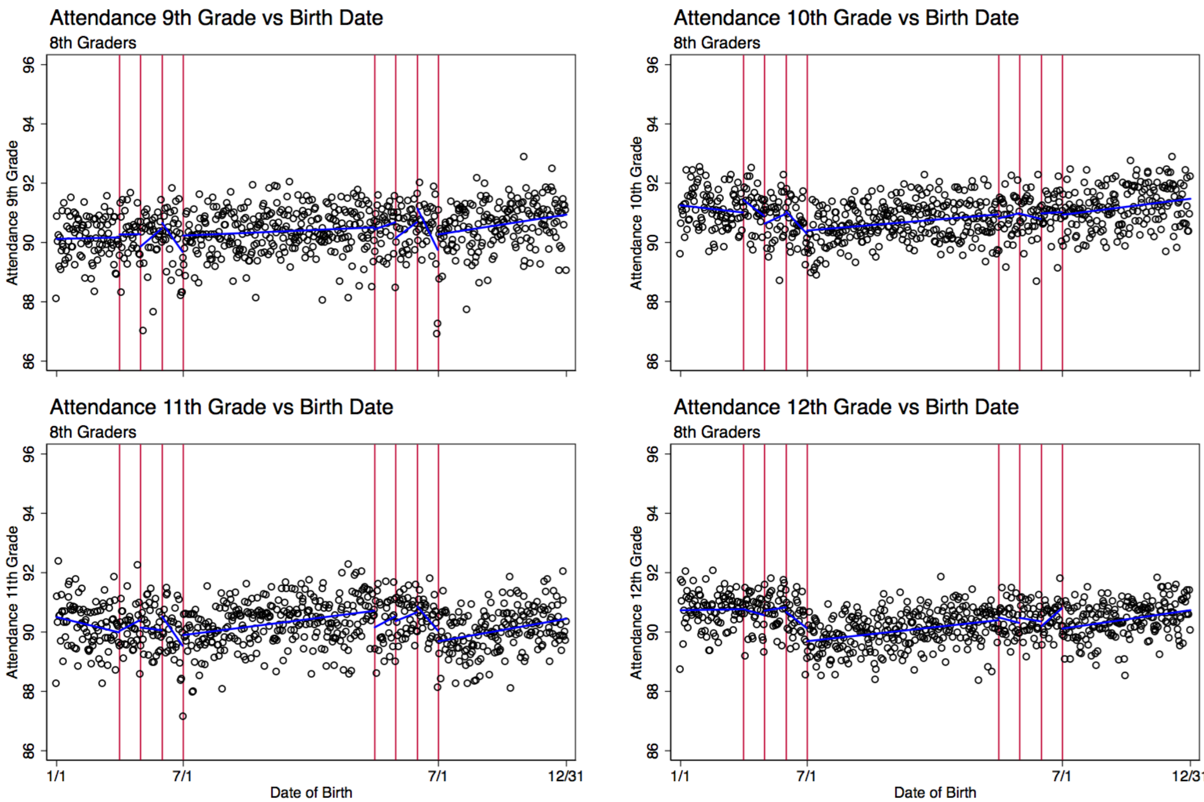
Notes: This Figures show how the respective outcomes behave when plotted against birth date. The dots plot the mean of each variable within birthday bins, and the solid line plots fitted values from local linear regressions on both sides of the birth date cutoff.

Figure 2.6: Outcomes during Primary School (1st to 8th grade): Attendance



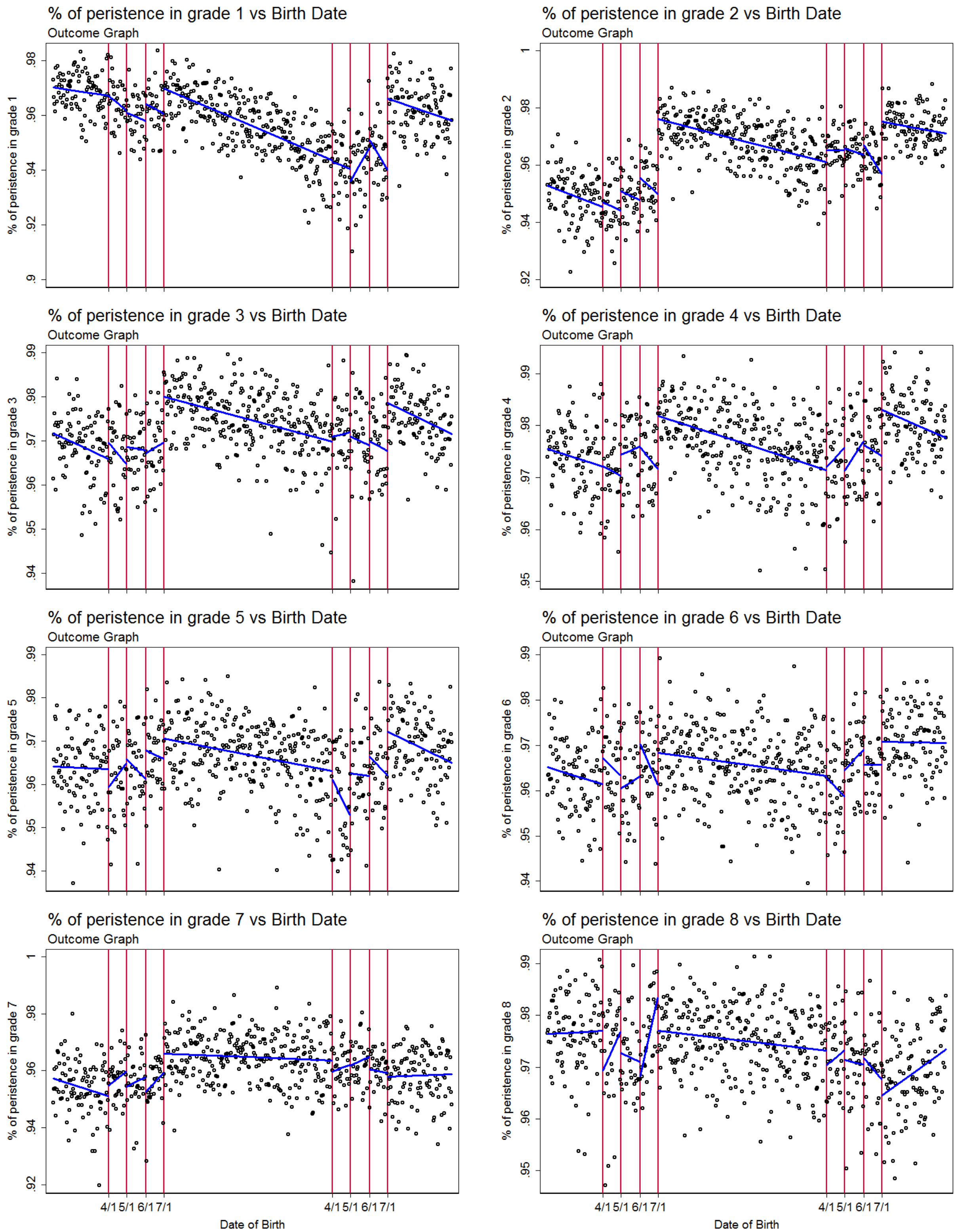
Notes: This Figures show how the respective outcomes behave when plotted against birth date. The dots plot the mean of each variable within birthday bins, and the solid line plots fitted values from local linear regressions on both sides of the birth date cutoff.

Figure 2.7: Outcomes during High School (until 12th grade): Attendance



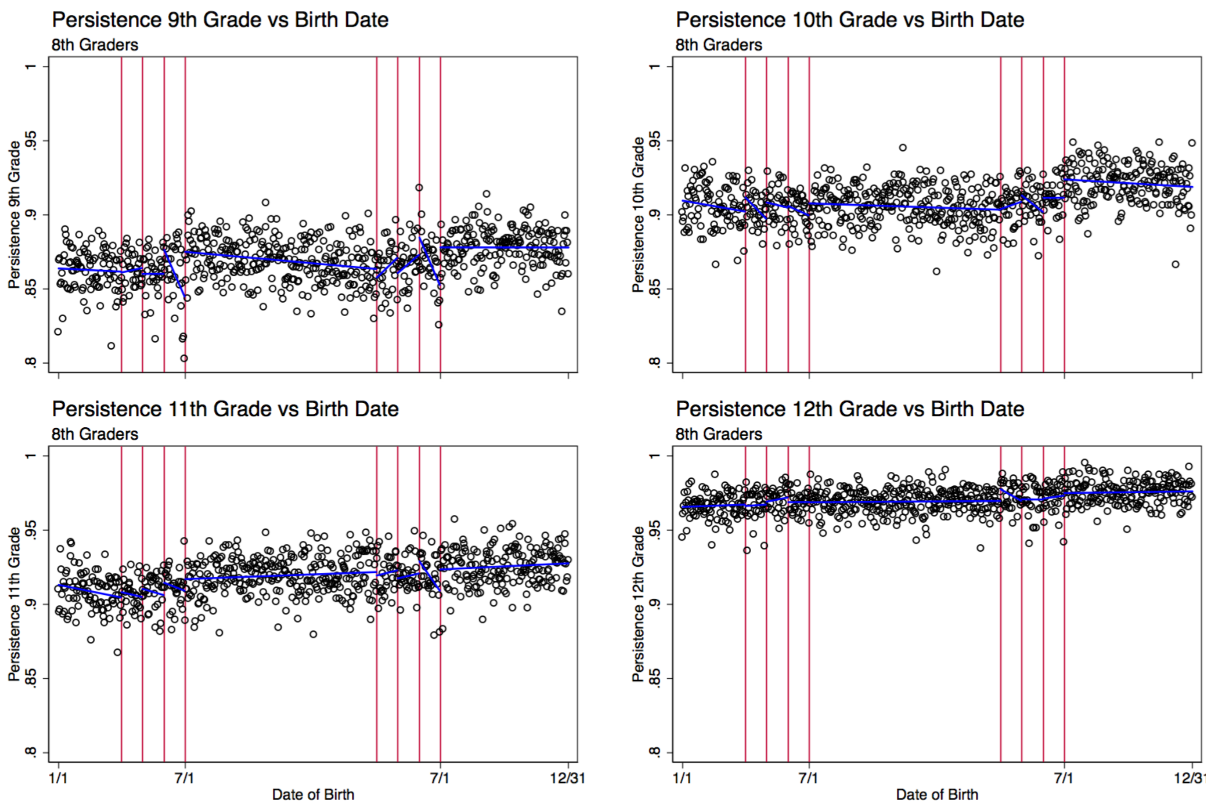
Notes: This Figures show how the respective outcomes behave when plotted against birth date. The dots plot the mean of each variable within birthday bins, and the solid line plots fitted values from local linear regressions on both sides of the birth date cutoff.

Figure 2.8: Outcomes during Primary School (1st to 8th grade): Persistence



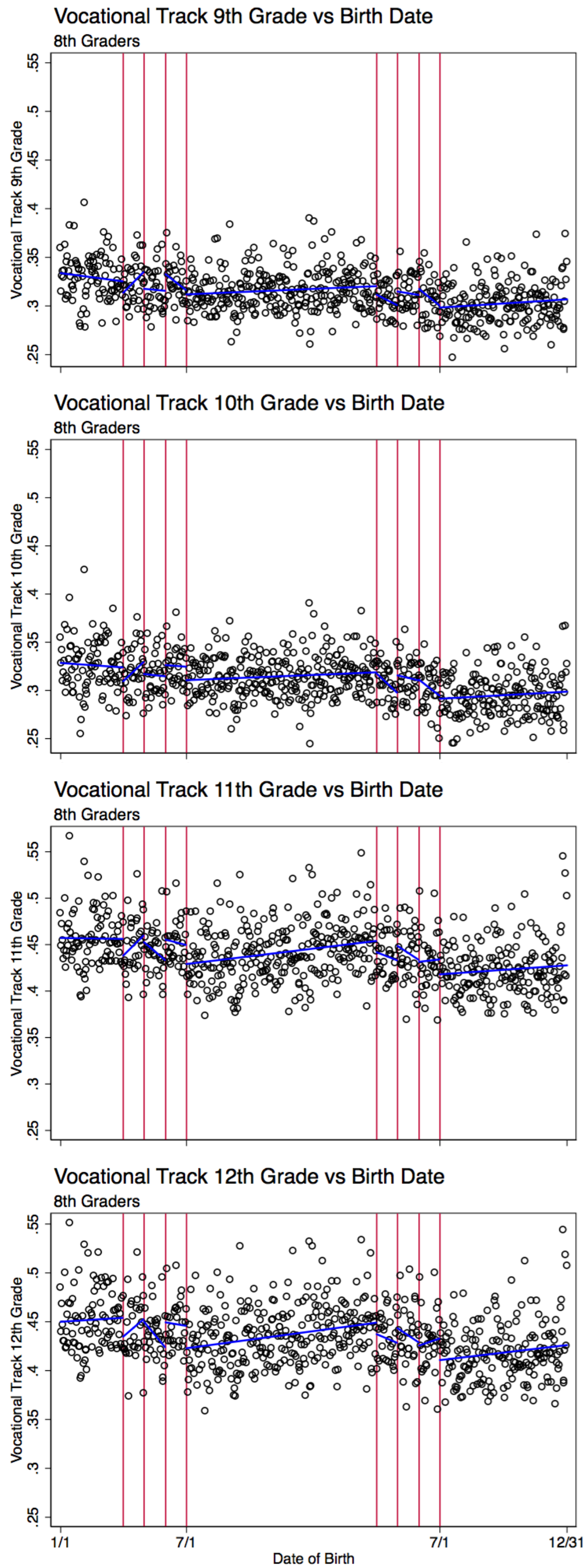
Notes: This Figures show how the respective outcomes behave when plotted against birth date. The dots plot the mean of each variable within birthday bins, and the solid line plots fitted values from local linear regressions on both sides of the birth date cutoff.

Figure 2.9: Outcomes during High School (until 12th grade): Persistence



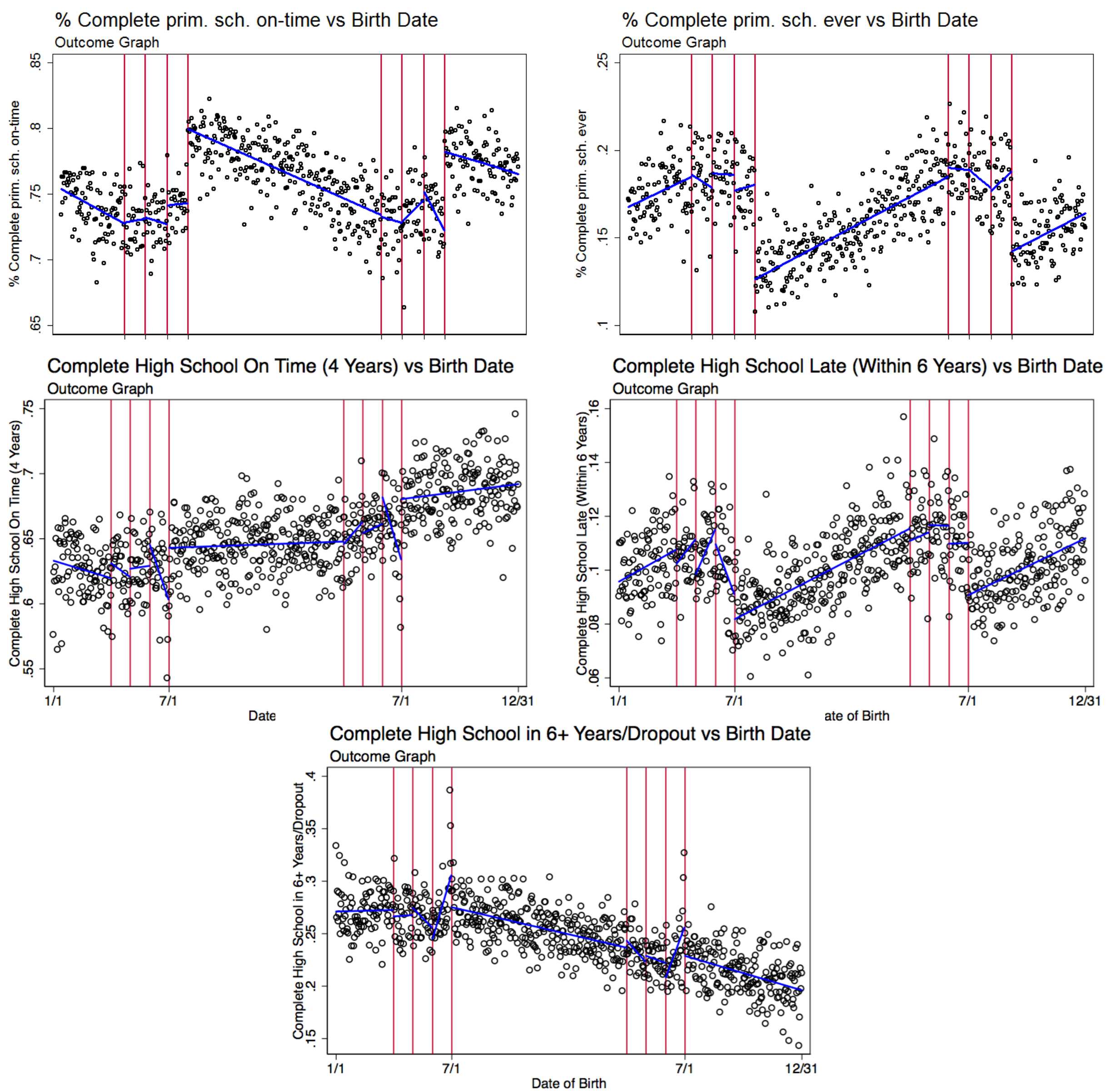
Notes: This Figures show how the respective outcomes behave when plotted against birth date. The dots plot the mean of each variable within birthday bins, and the solid line plots fitted values from local linear regressions on both sides of the birth date cutoff.

Figure 2.10: Outcomes during High School (until 12th grade): Vocational Track



Notes: This Figures show how the respective outcomes behave when plotted against birth date. The dots plot the mean of each variable within birthday bins, and the solid line plots fitted values from local linear regressions on both sides of the birth date cutoff.

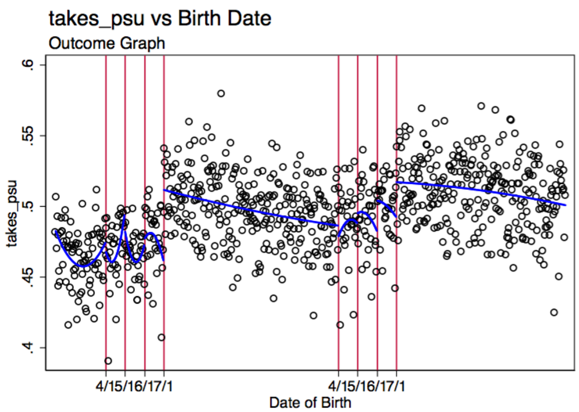
Figure 2.11: Attainment: Primary School and High School



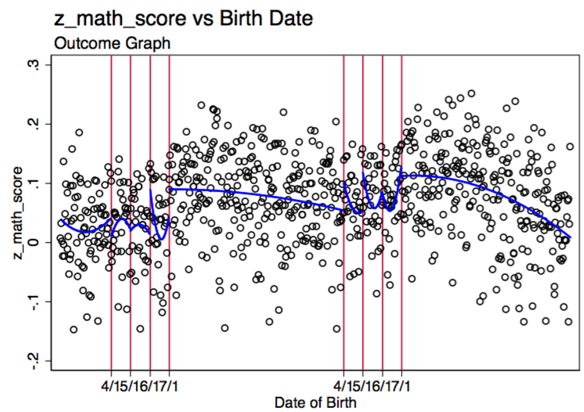
Notes: This Figures show how the respective outcomes behave when plotted against birth date. The dots plot the mean of each variable within birthday bins, and the solid line plots fitted values from local linear regressions on both sides of the birth date cutoff.

Figure 2.12: Transition to Postsecondary Education

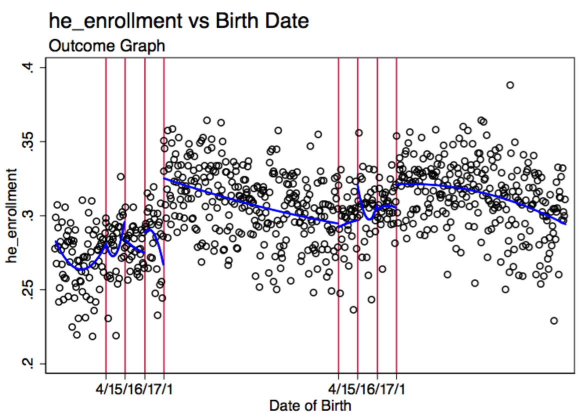
(a) Admission Test Take-up (timely)



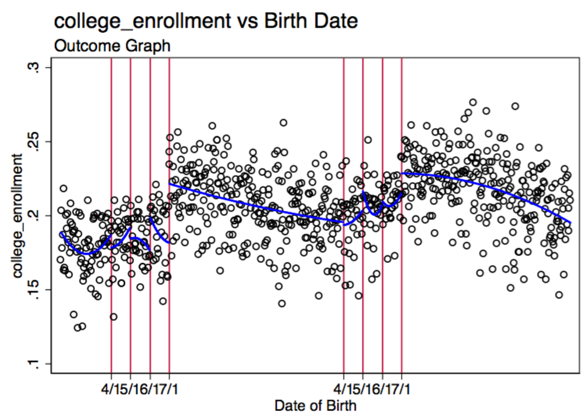
(b) PSU Z-Score (Math)



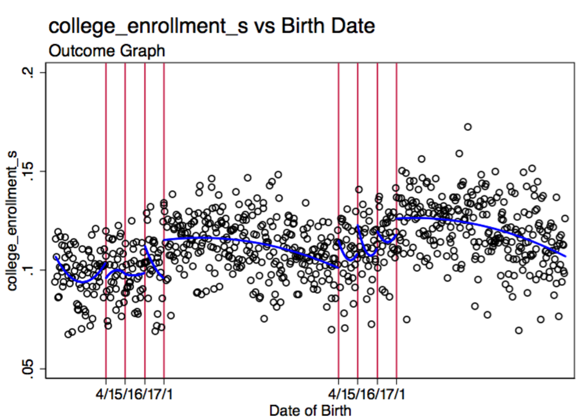
(c) Higher Education Enrollment



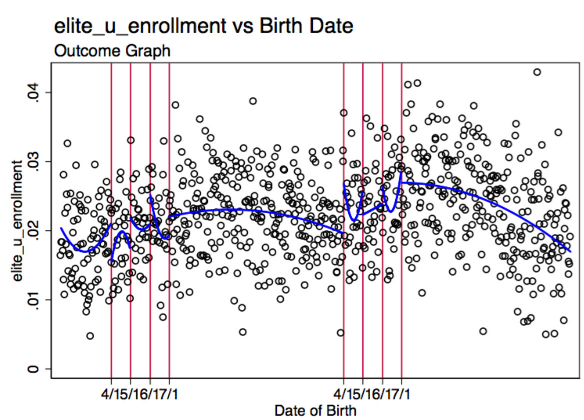
(d) College Enrollment (Any)



(e) College Enrollment (Selective)

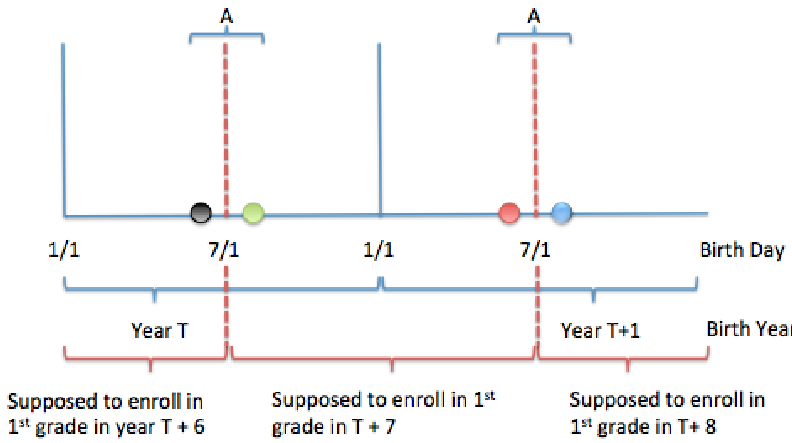


(f) College Enrollment (Elite)



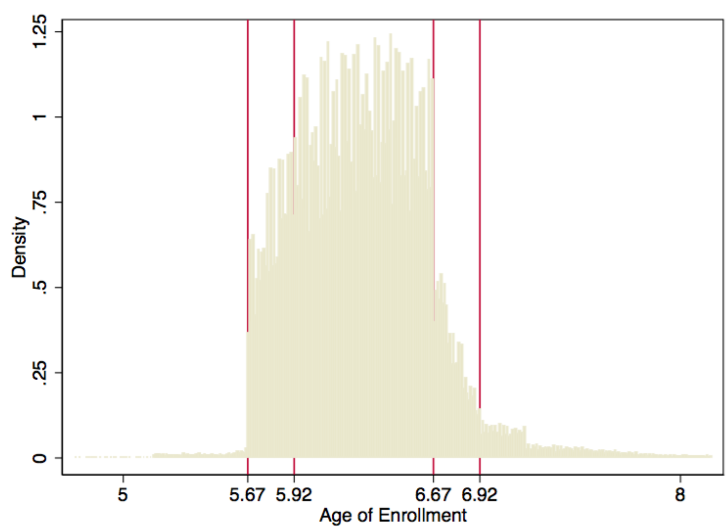
Notes: This Figures show how the respective outcomes behave when plotted against birth date. The dots plot the mean of each variable within birthday bins, and the solid line plots fitted values from local linear regressions on both sides of the birth date cutoff.

Figure 2.13: First Grade Enrollment Cutoffs



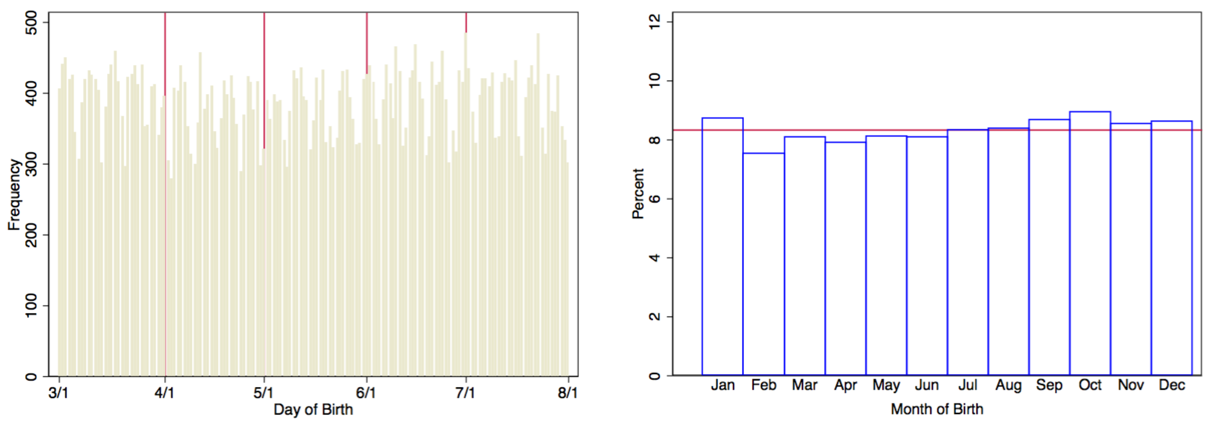
Notes: The Figure shows how the delayed enrollment works for three adjacent cohorts. The horizontal axis represents a time-line for birthdays over years T and T+1. In the middle of both years, the dashed line marks the cutoff date of July 1. Children in the vicinity defined by 'A,' were born only days apart, but those falling on the right side delayed enrollment by one year because they just missed the cutoff date. Therefore, those born before July 1 in year T are supposed to enroll in first grade in year T+6, i.e. when 6 years old. Those born on or after July 1 in year T delay their entrance to school and enroll with the next cohort, in year T+7, when they are 7 years old.

Figure 2.14: Enrollment Age in First Grade



Notes: The Figure shows the age of enrollment histogram. If all schools applied the July 1 cutoff, then ages 5.67 (first vertical line) to 6.67 (third vertical line) would be the common enrollment age interval in first grade. If April 1 cutoff is used then the interval would be between 5.92 (second vertical line) and 6.92 (fourth vertical line).

Figure 2.15: Birth Timing



Notes: We employ McCrary's (2008) density test to examine the existence of sorting in the vicinity of birth date threshold. This Figure presents a graphical illustration of this test, which shows that the density varies continuously over birth-dates with no significant discontinuity around the threshold.

Table 2.3: Reduced form estimates: 1st Graders 2002 - 2003

REDUCED FORM: 1st Graders 2002 - 2003

	First Stage	**Z-Score GPA**								**Test Scores**	
	Age in 1st grade	1st Grade	2nd Grade	3rd Grade	4th Grade	5th grade	6th grade	7th grade	8th grade	Math	Language
ITT	0.494***	0.305***	0.236***	0.220***	0.186***	0.134***	0.127***	0.089***	0.074***	0.195***	0.183***
Bandwidth +- 30 days	(0.009)	(0.019)	(0.019)	(0.020)	(0.020)	(0.021)	(0.021)	(0.022)	(0.022)	(0.021)	(0.021)
Local linear regression											
Observations	41578	41011	39105	37962	37015	35930	35023	33932	32632	34204	34183
ITT	0.482***	0.285***	0.225***	0.202***	0.173***	0.128***	0.109***	0.091***	0.090***	0.176***	0.169***
Bandwidth +- 82 days	(0.007)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Local linear regression											
Observations	112704	111201	105767	102562	100036	96874	94359	91456	87926	92417	92384

	First Stage	**Attendance**							
	Age in 1st grade	1st Grade	2nd Grade	3rd Grade	4th Grade	5th grade	6th grade	7th grade	8th grade
ITT	0.494***	0.769***	0.579***	-0.066	0.175	0.237	-0.167	-0.015	-0.133
Bandwidth +- 30 days	(0.009)	(0.246)	(0.173)	(0.155)	(0.155)	(0.163)	(0.155)	(0.194)	(0.188)
Local linear regression									
Observations	41578	41578	39382	38187	37216	36136	35191	34092	32787
ITT	0.482***	0.690***	0.346***	-0.066	-0.009	-0.005	-0.240**	-0.138	0.103
Bandwidth +- 82 days	(0.007)	(0.158)	(0.108)	(0.101)	(0.102)	(0.098)	(0.096)	(0.114)	(0.117)
Local linear regression									
Observations	112704	112704	106527	103173	100569	97391	94854	91832	88353

	First Stage	**Persistence**								**Timely graduation**	
	Age in 1st grade	1st Grade	2nd Grade	3rd Grade	4th Grade	5th grade	6th grade	7th grade	8th grade	Timely	Non timely
ITT	0.494***	0.007**	0.027***	0.009***	0.011***	0.005	0.007*	0.007*	-0.006**	0.056***	-0.052***
Bandwidth +- 30 days	(0.009)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)	(0.008)	(0.007)
Local linear regression											
Observations	41578	41578	39382	38187	37216	36136	35191	34092	32787	41578	41578
ITT	0.482***	0.007***	0.021***	0.010***	0.009***	0.002	0.002	0.009***	-0.000	0.050***	-0.048***
Bandwidth +- 82 days	(0.007)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.005)	(0.004)
Local linear regression											
Observations	112704	112704	106527	103173	100569	97391	94854	91832	88353	112704	112704

Notes: The table reports the reduced form estimates for the described set students. ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively. Robust standard errors clustered at the birth day level are in parentheses.

Table 2.4: LATE estimates: 1st Graders 2002 - 2003

LATE: 1st Graders 2002 - 2003

	Z-Score GPA								Test Scores	
	1st Grade	2nd Grade	3rd Grade	4th Grade	5th grade	6th grade	7th grade	8th grade	Math	Language
Age of enrollment	0.589***	0.492***	0.427***	0.365***	0.274***	0.232***	0.170***	0.182***	0.367***	0.358***
Bandwidth +- 30 days	(0.021)	(0.021)	(0.022)	(0.023)	(0.023)	(0.023)	(0.024)	(0.025)	(0.024)	(0.024)
Local linear regression										
Observations	41011	39105	37962	37015	35930	35023	33932	32632	34204	34183
Age of enrollment	0.602***	0.507***	0.430***	0.385***	0.288***	0.228***	0.164***	0.169***	0.378***	0.361***
Bandwidth +- 82 days	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)	(0.015)	(0.015)
Local linear regression										
Observations	111201	105767	102562	100036	96874	94359	91456	87926	92417	92384

	Attendance							
	1st Grade	2nd Grade	3rd Grade	4th Grade	5th grade	6th grade	7th grade	8th grade
Age of enrollment	1.311***	0.674***	-0.121	-0.026	-0.101	-0.744***	-0.537**	0.411*
Bandwidth +- 30 days	(0.292)	(0.204)	(0.193)	(0.191)	(0.187)	(0.176)	(0.212)	(0.219)
Local linear regression								
Observations	41578	39382	38187	37216	36136	35191	34092	32787
Age of enrollment	1.845***	0.816***	-0.006	0.067	-0.062	-0.720***	-0.609***	0.570***
Bandwidth +- 82 days	(0.197)	(0.142)	(0.119)	(0.116)	(0.117)	(0.109)	(0.119)	(0.135)
Local linear regression								
Observations	112704	106527	103173	100569	97391	94854	91832	88353

	Persistence								Timely graduation	
	1st Grade	2nd Grade	3rd Grade	4th Grade	5th grade	6th grade	7th grade	8th grade	Timely	Non timely
Age of enrollment	0.011***	0.048***	0.023***	0.017***	0.006	0.005	0.019***	0.002	0.108***	-0.101***
Bandwidth +- 30 days	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.005)	(0.004)	(0.009)	(0.008)
Local linear regression										
Observations	41578	39382	38187	37216	36136	35191	34092	32787	41578	41578
Age of enrollment	0.010***	0.049***	0.022***	0.015***	0.009***	0.009***	0.019***	0.007***	0.115***	-0.098***
Bandwidth +- 82 days	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.005)	(0.005)
Local linear regression										
Observations	112704	106527	103173	100569	97391	94854	91832	88353	112704	112704

Notes: The table reports the LATE estimates for the described set students. ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively. Robust standard errors clustered at the birth day level are in parentheses.

Table 2.5: Reduced form estimates: 1st Graders 2003 - 2004

REDUCED FORM: 1st Graders 2003 - 2004

	First Stage	**Z-Score GPA**								**Test Scores**	
	Age in 1st grade	1st Grade	2nd Grade	3rd Grade	4th Grade	5th grade	6th grade	7th grade	8th grade	Math	Language
ITT	0.471***	0.305***	0.242***	0.230***	0.205***	0.153***	0.112***	0.067***	0.086***	0.201***	0.205***
Bandwidth +- 30 days	(0.009)	(0.019)	(0.020)	(0.020)	(0.020)	(0.021)	(0.020)	(0.021)	(0.022)	(0.022)	(0.022)
Local linear regression											
Observations	42029	41445	39442	38213	37175	36356	35330	34219	32812	34294	34300
ITT	0.443***	0.269***	0.221***	0.206***	0.181***	0.140***	0.096***	0.062***	0.071***	0.153***	0.170***
Bandwidth +- 82 days	(0.007)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Local linear regression											
Observations	111931	110337	104866	101554	98752	96475	93469	90599	86895	91218	91139

	First Stage	**Attendance**							
	Age in 1st grade	1st Grade	2nd Grade	3rd Grade	4th Grade	5th grade	6th grade	7th grade	8th grade
ITT	0.471***	0.900***	0.312	0.072	0.341**	-0.179	-0.295*	-0.176	-1.151***
Bandwidth +- 30 days	(0.009)	(0.246)	(0.203)	(0.173)	(0.169)	(0.168)	(0.158)	(0.177)	(0.237)
Local linear regression									
Observations	42029	42029	39708	38435	37375	36546	35506	34360	32990
ITT	0.443***	0.539***	0.251**	0.066	0.067	-0.284***	-0.288***	0.005	-0.744***
Bandwidth +- 82 days	(0.007)	(0.151)	(0.117)	(0.111)	(0.098)	(0.094)	(0.096)	(0.111)	(0.144)
Local linear regression									
Observations	111931	111931	105579	102157	99317	96990	93961	90958	87390

	First Stage	**Persistence**								**Timely graduation**	
	Age in 1st grade	1st Grade	2nd Grade	3rd Grade	4th Grade	5th grade	6th grade	7th grade	8th grade	Timely	Non timely
ITT	0.471***	0.031***	0.017***	0.010***	0.010***	0.010**	0.008**	-0.004	-0.004	0.064***	-0.047***
Bandwidth +- 30 days	(0.009)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.008)	(0.007)
Local linear regression											
Observations	42029	42029	39708	38435	37375	36546	35506	34360	32990	42029	42029
ITT	0.443***	0.019***	0.014***	0.011***	0.007***	0.006***	0.004*	-0.003	-0.003	0.045***	-0.039***
Bandwidth +- 82 days	(0.007)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.005)	(0.004)
Local linear regression											
Observations	111931	111931	105579	102157	99317	96990	93961	90958	87390	111931	111931

Notes: The table reports the reduced form estimates for the described set students. ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively. Robust standard errors clustered at the birth day level are in parentheses.

Table 2.6: LATE estimates: 1st Graders 2003 - 2004

LATE: 1st Graders 2002 - 2003

	Z-Score GPA								**Test Scores**	
	1st Grade	2nd Grade	3rd Grade	4th Grade	5th grade	6th grade	7th grade	8th grade	Math	Language
Age of enrollment	0.601***	0.492***	0.453***	0.405***	0.313***	0.221***	0.136***	0.155***	0.337***	0.385***
Bandwidth +- 30 days	(0.022)	(0.023)	(0.023)	(0.024)	(0.024)	(0.025)	(0.025)	(0.026)	(0.026)	(0.026)
Local linear regression										
Observations	41445	39442	38213	37175	36356	35330	34219	32812	34294	34300
Age of enrollment	0.631***	0.514***	0.448***	0.417***	0.309***	0.223***	0.129***	0.133***	0.343***	0.369***
Bandwidth +- 82 days	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)	(0.014)	(0.014)	(0.015)	(0.016)	(0.016)
Local linear regression										
Observations	110337	104866	101554	98752	96475	93469	90599	86895	91218	91139
	Attendance									
	1st Grade	2nd Grade	3rd Grade	4th Grade	5th grade	6th grade	7th grade	8th grade		
Age of enrollment	1.343***	0.533**	0.153	0.126	-0.665***	-0.516***	0.328	-1.878***		
Bandwidth +- 30 days	(0.280)	(0.218)	(0.208)	(0.193)	(0.187)	(0.191)	(0.222)	(0.282)		
Local linear regression										
Observations	42029	39708	38435	37375	36546	35506	34360	32990		
Age of enrollment	1.450***	0.232	0.117	0.163	-0.733***	-0.511***	0.459***	-2.481***		
Bandwidth +- 82 days	(0.209)	(0.151)	(0.140)	(0.115)	(0.113)	(0.114)	(0.128)	(0.172)		
Local linear regression										
Observations	111931	105579	102157	99317	96990	93961	90958	87390		
	Persistence								**Timely graduation**	
	1st Grade	2nd Grade	3rd Grade	4th Grade	5th grade	6th grade	7th grade	8th grade	Timely	Non timely
Age of enrollment	0.044***	0.030***	0.021***	0.018***	0.015***	0.014***	-0.007	-0.009*	0.100***	-0.086***
Bandwidth +- 30 days	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.010)	(0.008)
Local linear regression										
Observations	42029	39708	38435	37375	36546	35506	34360	32990	42029	42029
Age of enrollment	0.045***	0.020***	0.014***	0.015***	0.020***	0.014***	-0.007**	-0.009***	0.091***	-0.078***
Bandwidth +- 82 days	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.006)	(0.005)
Local linear regression										
Observations	111931	105579	102157	99317	96990	93961	90958	87390	111931	111931

Notes: The table reports the LATE estimates for the described set students. ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively. Robust standard errors clustered at the birth day level are in parentheses.

Table 2.7: Reduced form estimates: 8th Graders 2003 - 2004

	First Stage	**Z-Score GPA**				
	Age in 1st grade	8th grade	9th Grade	10th Grade	11th Grade	12th Grade
ITT	0.341***	0.157***	0.123***	0.082***	0.036	0.038
Bandwidth +- 30 days	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Local linear regression						
Observations	38,524	38,302	35,542	31,468	28,528	26,417
ITT	0.384***	0.132***	0.091***	0.066***	0.027*	0.030*
Bandwidth +- 70 days	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Local linear regression						
Observations	90,411	89,933	83,382	73,956	66,972	62,051

	First Stage	**Attendance**				
	Age in 1st grade	8th grade	9th Grade	10th Grade	11th Grade	12th Grade
ITT	0.341***	-0.131	-0.064	-0.560*	0.127	-0.514*
Bandwidth +- 30 days	(0.01)	(0.23)	(0.37)	(0.34)	(0.40)	(0.31)
Local linear regression						
Observations	38,524	38,524	36,521	32,138	29,283	26,730
ITT	0.384***	-0.158	-0.289	-0.522**	-0.2	-0.691***
Bandwidth +- 70 days	(0.01)	(0.14)	(0.23)	(0.22)	(0.25)	(0.20)
Local linear regression						
Observations	90,411	90,411	85,658	75,506	68,664	62,802

	First Stage	**Persistence**					**Timely graduation High School**		
	Age in 1st grade	8th grade	9th Grade	10th Grade	11th Grade	12th Grade	Timely	Non timely	Never
ITT	0.341***	-0.003	0.025***	-0.002	0.004	0.002	0.022**	-0.003	-0.019**
Bandwidth +- 30 days	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
Local linear regression									
Observations	38,524	38,524	36,521	32,138	29,283	26,730	38,524	38,524	38,524
ITT	0.384***	0.001	0.016***	0.001	-0.002	-0.001	0.011*	-0.012***	0.001
Bandwidth +- 70 days	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)
Local linear regression									
Observations	90,411	90,411	85,658	75,506	68,664	62,802	90,411	90,411	90,411

Notes: The table reports the reduced form estimates for the described set students. ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively. Robust standard errors clustered at the birth day level are in parentheses.

Table 2.8: Reduced form estimates: 8th Graders 2004 - 2005

	First Stage	**Z-Score GPA**				
	Age in 1st grade	8th grade	9th Grade	10th Grade	11th Grade	12th Grade
ITT	0.342***	0.135***	0.089***	0.074***	0.054**	0.005
Bandwidth +- 30 days	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Local linear regression						
Observations	38,220	38,005	35,089	30,746	28,153	26,016
ITT	0.376***	0.114***	0.067***	0.065***	0.055***	0.021
Bandwidth +- 70 days	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Local linear regression						
Observations	89,745	89,218	82,413	72,207	66,038	61,235

	First Stage	**Attendance**				
	Age in 1st grade	8th grade	9th Grade	10th Grade	11th Grade	12th Grade
ITT	0.342***	0.386*	-0.218	-0.088	-0.15	-0.651**
Bandwidth +- 30 days	(0.01)	(0.22)	(0.40)	(0.36)	(0.38)	(0.33)
Local linear regression						
Observations	38,220	38,220	36,225	31,389	28,755	26,330
ITT	0.376***	0.137	-0.366	-0.062	-0.651***	-0.454**
Bandwidth +- 70 days	(0.01)	(0.14)	(0.25)	(0.23)	(0.24)	(0.20)
Local linear regression						
Observations	89,745	89,745	85,084	73,767	67,586	61,940

	First Stage	**Persistence**					**Timely graduation High School**		
	Age in 1st grade	8th grade	9th Grade	10th Grade	11th Grade	12th Grade	Timely	Non timely	Never
ITT	0.342***	0.004	0.013*	0.015**	0.013*	-0.002	0.027***	-0.018***	-0.009
Bandwidth +- 30 days	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
Local linear regression									
Observations	38,220	38,220	36,225	31,389	28,755	26,330	38,220	38,220	38,220
ITT	0.376***	0.007***	0.004	0.011***	0.006	0.002	0.017***	-0.019***	0.002
Bandwidth +- 70 days	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)
Local linear regression									
Observations	89,745	89,745	85,084	73,767	67,586	61,940	89,745	89,745	89,745

Notes: The table reports the reduced form estimates for the described set students. ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively. Robust standard errors clustered at the birth day level are in parentheses.

Table 2.9: LATE estimates: 8th Graders 2003 - 2004

	Z-Score GPA				
	8th grade	9th Grade	10th Grade	11th Grade	12th Grade
Age of enrollment	0.331***	0.223***	0.151***	0.083***	0.072**
Bandwidth +- 30 days	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Local linear regression					
Observations	38,302	35,542	31,468	28,528	26,417
Age of enrollment	0.353***	0.243***	0.163***	0.114***	0.079***
Bandwidth +- 70 days	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Local linear regression					
Observations	89,933	83,382	73,956	66,972	62,051

	Attendance				
	8th grade	9th Grade	10th Grade	11th Grade	12th Grade
Age of enrollment	-0.241	-0.144	-1.155***	-0.48	-1.637***
Bandwidth +- 30 days	(0.34)	(0.46)	(0.41)	(0.47)	(0.37)
Local linear regression					
Observations	38,524	36,521	32,138	29,283	26,730
Age of enrollment	0.223	0.497	-0.977***	-0.429	-2.085***
Bandwidth +- 70 days	(0.31)	(0.31)	(0.27)	(0.31)	(0.25)
Local linear regression					
Observations	90,411	85,658	75,506	68,664	62,802

	Persistence					**Timely graduation High School**		
	8th grade	9th Grade	10th Grade	11th Grade	12th Grade	Timely	Non timely	Never
Age of enrollment	0.007	0.037***	0.004	0.002	0.004	0.043***	-0.038***	-0.006
Bandwidth +- 30 days	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Local linear regression								
Observations	38,524	36,521	32,138	29,283	26,730	38,524	38,524	38,524
Age of enrollment	0.008**	0.043***	0.006	0.019***	0	0.065***	-0.046***	-0.019**
Bandwidth +- 70 days	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
Local linear regression								
Observations	90,411	85,658	75,506	68,664	62,802	90,411	90,411	90,411

Notes: The table reports the LATE estimates for the described set students. ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively. Robust standard errors clustered at the birth day level are in parentheses.

Table 2.10: LATE estimates: 8th Graders 2004 - 2005

			Z-Score GPA		
	8th grade	9th Grade	10th Grade	11th Grade	12th Grade
Age of enrollment	0.307***	0.173***	0.170***	0.138***	0.065**
Bandwidth +- 30 days	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Local linear regression					
Observations	38,005	35,089	30,746	28,153	26,016
Age of enrollment	0.346***	0.201***	0.171***	0.128***	0.071***
Bandwidth +- 70 days	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Local linear regression					
Observations	89,218	82,413	72,207	66,038	61,235

			Attendance		
	8th grade	9th Grade	10th Grade	11th Grade	12th Grade
Age of enrollment	0.121	-0.964**	-0.118	-1.679***	-0.770**
Bandwidth +- 30 days	(0.27)	(0.49)	(0.44)	(0.46)	(0.39)
Local linear regression					
Observations	38,220	36,225	31,389	28,755	26,330
Age of enrollment	0.235	-0.475	0.094	-1.668***	-0.405
Bandwidth +- 70 days	(0.19)	(0.33)	(0.29)	(0.31)	(0.25)
Local linear regression					
Observations	89,745	85,084	73,767	67,586	61,940

			Persistence			**Timely graduation High School**		
	8th grade	9th Grade	10th Grade	11th Grade	12th Grade	Timely	Non timely	Never
Age of enrollment	0.014***	0.009	0.028***	0.008	0.007	0.040***	-0.046***	0.006
Bandwidth +- 30 days	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Local linear regression								
Observations	38,220	36,225	31,389	28,755	26,330	38,220	38,220	38,220
Age of enrollment	0.008***	0.022***	0.031***	0.010*	0.010***	0.055***	-0.050***	-0.005
Bandwidth +- 70 days	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
Local linear regression								
Observations	89,745	85,084	73,767	67,586	61,940	89,745	89,745	89,745

Notes: The table reports the LATE estimates for the described set students. ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively. Robust standard errors clustered at the birth day level are in parentheses.

CHAPTER 3

RESCUING LOW-INCOME, HIGH ABILITY STUDENTS: UNIVERSITY ACCESS, VALIDITY AND MISMATCH

3.1 Introduction

In many countries, university admissions mainly depend on high-stakes tests under the assumption that their scores provide an adequate measure of post-secondary promise. Assessing whether admission factors accurately predict future success remains a challenging task as the performance of those who were not admitted is missing from the data (Rothstein 2004). Moreover, test scores are typically highly correlated with socioeconomic status (Zwick and Green 2007), particularly in less developed countries, and do not adequately capture skills that predict success in life (Heckman and Kautz 2012). Taken together, these results suggest that test-score-based admissions may perpetuate education and income immobility by rejecting applicants – principally low-income students – who would have otherwise succeeded academically, despite lower test scores.

This paper evaluates whether admission standards screen out highly able but low socioeconomic status applicants, and whether admission policies could 'rescue' this sort of prospective student. To do so, I use information from an affirmative action (AA) policy implemented in Chile targeting top students from low-income households with scores below admission cutoffs. One of the major advantages in using the Chilean score-based data is that it provides the complete information used in determining the applicants' admission, which I use to explicitly address key questions regarding effects on access and academic outcomes.

As opposed to its typical use, the term 'affirmative action' in this context is not based on race or ethnicity, but on socioeconomic status (SES).[1] In the Chilean nationwide and

1. Race-based AA policies are diminishing in the U.S. and alternatives based on income or high school rank are being proposed. See e.g., the Supreme Court's decision in Fisher v. University of Texas (2013), Antonovics and Backes (2014), and Blume and Long (2014). For alternative AA policies, see Cashin (2014), Kahlenberg (2014), White House (2014) and Long (2015). The Chilean case provides an opportunity to

centralized admission system, prospective students apply upon entry to a set of specific university-major combinations at flagship institutions[2] with their high school GPA and admission exam scores. The system places great value on these scores ($\sim$ 80%), which are strongly correlated with SES (OECD 2010). Thus, economically disadvantaged students may not get admission even after earning the highest GPA at their high schools. The AA policy gave extra slots in admissions to low-income applicants who were among the best students (top 5% of GPA) from their high schools but whose test scores were not high enough to guarantee admission into flagship universities.

I gathered rich administrative records from different sources following students from their high-school years through their final years at university. In particular, I use information on the individuals' high school GPA, high school characteristics, admission exam scores, preferences for university-major combinations, demographics, first year GPA and higher education enrollment.

I examine whether the admission system would have rejected students with good academic futures by comparing differences in a host of academic outcomes - first year GPA, dropout rates and year-to-year persistence– between the AA beneficiaries and the group of regularly enrolled students that would have been displaced by them had the policy not consisted in extra slots. This analysis contributes to the *validity* literature (Rothstein 2004), that attempts to predict the performance of students based on admission factors and it is policy relevant for university officers in optimizing their admission schemes. Students with scores below the major cutoff are denied enrollment under the assumption that if admitted, they would perform worse than those above the threshold. Typically this statement is not testable, but the AA policy design allows to do so. The results show that, despite their lower scores AA students perform on average at least as well as the pool of regularly enrolled students in the

evaluate an AA policy designed in this latter spirit, to some extent resembling the 'percent plans' implemented in the U.S., but including a SES component.

2. The 25 flagship universities, also known as 'traditional universities,' represent the core set of universities in Chile. Historically, they have represented the largest percentage of the country's university enrollment.

same majors. Moreover, when comparing students closer to the admissions cutoff, I find that AA students attain sizable higher persistence rates than regularly enrolled students. This latter differential increased from 6 points during their freshman year, to up to 10 percentage points in their senior year.

These results contradict the score-based admission system prediction explicitly stated in the admissions official documents (DEMRE 2008). One explanation is that student ability is multidimensional and test scores measure only some dimensions of skills. Following a growing literature relating different set of skills and adult outcomes such as educational attainment (Heckman, Stixrud, and Urzua 2006; Almlund et al. 2011; and Heckman and Kautz 2012, 2014a, 2014b) I estimate a factor model using test scores and additional non-test-score measures of skill that reflect behaviors that go beyond pure cognition, like showing up for class (Kautz and Zanoni 2015). I find that AA students have higher non-test-score skills than their regularly enrolled peers. However, AA students have lower test-score skills, lower parental schooling and also attended lower quality high-schools. Once controlling for the non-test-score skill factor in estimating university persistence, the advantage favoring AA students vanishes. This evidence suggest that AA students are equipped with a set of skills not measured by test scores that compensate for their socioeconomic and high-school quality disadvantages, and enable them to outperform their peers in their same majors.

Another important dimension of the policy is how it affected students at the individual level. I ask two counterfactual questions on this respect. The first is how the policy affected enrollment[3] and the majors at which students ended up, if any. The second is how AA-eligible students perform in more selective majors compared to how they would have fared in less selective majors in absence of the policy, conditional on enrollment.

Regarding the first question, I find that two out of three eligible applicants took advantage

3. Most of the previous work on this area in the U.S. studies minority enrollment rates after an affirmative action ban has passed, as has been the case in many states (see Long and Tienda (2008), Hinrichs (2012) and Backes (2012), to name a few). See Kapor (2015) and Arcidiacono and Lovenheim (2015) for excellent surveys on this topic in the U.S.

of the policy. As a result of this take-up, nearly five hundred students –who would have been otherwise rejected by their preferred university-major combination– enrolled. I use their preferences for university-major pairs to identify which would have been their next-best alternative in the absence of the policy. Recent work by Kirkeboen, Leuven, and Mogstad (2014) highlights the importance of next-best alternatives to understand the effect of a policy that changes the applicant's field of study. Daugherty, Martorell and McFarlin (2014) found that a similar policy implemented in Texas simply re-sorted students among similar-quality institutions. However, my enrollment counterfactuals indicate that 84% of the beneficiaries enrolled in more selective majors, with higher expected wages and at more prestigious universities than they would have been able to without the policy.[4] Besides this 'intensive margin' effect, I also document that the policy had an 'extensive margin' effect for the remaining 16% of beneficiaries that would not have been admitted to any major at flagship universities had the policy not existed.

In addressing the second question I assess a version of the *mismatch hypothesis* (Sander 2004, and Sander and Taylor 2012) that states that AA-eligible would have fared better in other – less selective – majors. Using a regression discontinuity on the class rank condition, I find that eligible AA students would not have attained significantly higher persistence rates in less selective majors. This result is in line with the findings by papers that used similar discontinuity methods for the Texas case, like Niu and Tienda (2010), Fletcher and Mayer (2013), Daugherty, et al. (2014) and for Israel (Alon and Malamud 2014).

In sum, the empirical findings show that the AA policy contributed to equity by introducing low-income students to school-major combinations that would otherwise have rejected them.[5] Moreover, AA students exhibit good academic performance relative to their

4. I interpret this result as an indicator that AA students enroll in higher quality university-major combinations, and not necessarily as higher economic returns. I can not estimate these returns because wages for the cohort under analysis are not available yet. Hastings, Neilson and Zimmerman (2013) and Reyes, Rodriguez and Urzua (2013) examined economic returns for older cohorts in Chile. Even though both articles find that returns are heterogeneous, they estimate large, positive returns to the more selective university-majors combinations.

5. Hoxby and Avery (2013) show that in the U.S. those who are economically disadvantaged choose

peers, which indicates that the policy increased the efficiency of the system. This set of results suggests that the university admission systems could be improved by considering additional measures of skills using administrative data that are readily available from high-school records. Overall, there may be an opportunity to move forward on both equity and efficiency in higher education admissions.

The remainder of the paper is organized as follows. Section 3.2 provides background on the Chilean educational system and the special admission program. Section 3.3 outlines my empirical strategy, while Section 3.4 describes the data used. Section 3.5 presents and discusses the results. Section 3.6 concludes.

3.2 Background

3.2.1 The Chilean Context

Access to higher education in Chile has greatly expanded over the past 25 years. The number of students enrolled in postsecondary institutions has increased five fold, from less than 250,000 in 1990 to 1.2 million in 2014 (Mineduc 2014). This amount represents 7% of the total population (17 million), which is particularly high considering other Latin American countries (Unesco-OECD 2010) and is similar in magnitude to the U.S.[6]

However, despite progress in making higher education more accessible, challenges still remain. The country displays a high correlation between international large scale assessments[7] and socioeconomic status comparatively speaking. The variance of the Program for

colleges at the bottom of the quality distribution even though they would have been a better match for a school in the middle of the quality distribution. In the context of this paper, it seems that the admission system itself drives poorer university choices for the particular set of AA students.

6. As a benchmark, between 1991 and 2011 enrollment in U.S. degree-granting institutions increased from 14.3 million to 21.0 million (US Department of Education 2013), which represents 7% of the overall population. Besides the increasing enrollment in higher education Chile has also reached almost universal levels of educational coverage in primary (99%) and secondary school (92%), again well above Latin American countries (Unesco-OECD 2010).

7. According to PISA results (OECD 2009, 2012), Chile ranks among the best in Latin America. However, results are not as good when compared with other countries with similar GDP per capita. Chile stands below

International Student Assessment (PISA) scores explained by SES background was greater in Chile than any other Organization for Economic Cooperation and Development (OECD) member countries in 2010 (OECD 2010), except for Hungary. This strong correlation also appears when considering test scores from the national standardized examinations applied during primary and secondary school according to reports by the OECD and The World Bank (2009).

Regarding university enrollment, most of its increase has been due to the creation of private non-selective and mid-to-low quality universities, typically serving low-income students. Access to flagship universities, which are supported financially by the state and include the country's most selective and prestigious institutions, has increased at a much lower rate and remains characterized by persistent socioeconomic influence. Applicants from poorer households are less likely to achieve the high admission exam scores that give access to the best universities and therefore matriculate at much lower rates than do higher-income students (OECD and The World Bank 2009).

From the perspective of the admission system efficiency, the correlation between test scores and student socioeconomic status would not be a problem as long as it does not overlook prospective students with high academic potential. However, international reports (ETS 2005, Pearson 2013) evaluating the PSU scores have found prediction validity indices smaller than those reported internationally (Kobrin et al. 2008). Hence, the admission system faces a double challenge. On the one hand it confronts an equity challenge given the nature of the socioeconomic influence described above. On the other hand there is an efficiency challenge; the admission system does not want to reject students with good academic projections.

---

countries like Poland, Croatia, Lithuania and Turkey (PISA 2012).

3.2.2 The University Admission System

After graduating from high school, prospective students take a national admission test (Prueba de Seleccion Universitaria or PSU), similar to the SAT in the U.S. Test takers complete exams in mathematics and language, whose scores are standardized to have a mean of 500 points and a standard deviation of 110. As in most of the OECD countries (Kirkeboen, Leuven and Mogstad 2014), Chilean students choose a specific major upon entry, which differs from other systems in which the choice is made later in college (like in the U.S.).[8] Prospective students apply to up to eight university-major combinations ranked based on personal preference, via a centralized process. To simplify language, from now on I will call the university-major combination just 'major' unless otherwise noted. The students' application score for each preference is a major-specific weighted average of their PSU test scores ($\sim$ 80%) and their high school GPA ($\sim$ 20%). This score is assumed to predict a student's academic performance during her university studies as explicitly declared in the admissions official documents (DEMRE 2008). The allocation process ranks all the students applying to any given major using their respective scores.

This process mimics a deferred acceptance algorithm (as described in Abdulkadiroglu, Angrist and Pathak 2014), in which students apply to their first choice in the first round, and each major rejects the students with lowest application score in excess of its capacity. In the next round (r), students rejected in the previous round (r-1) apply to their next most preferred major (if any). Each major considers these students and provisionally admitted students from the previous round combined, again rejecting the lowest-scoring students in excess of capacity and producing a new provisional admit list. This procedure defines a score cutoff for each major. The algorithm ends when either every student is matched to one major or every major has rejected every unmatched student. If a student is matched into more than one program, she is placed in her highest preference and is eliminated from

8. The university system in Chile is organized around a structured major (i.e. engineering, medicine, business etc.), with a strong core curriculum. Students have very little opportunity to switch from one curriculum to another as opposed to common practice in the U.S., where there is much more flexibility.

all other rankings. Hence there are no incentives to place less preferred majors before more preferred ones and it is optimal to reveal the true preferences for majors (Bordon and Fu 2015).[9]

In practice, applicants receive the admission results containing their application and cutoff scores (unknown at the moment of application) for each of their listed preferences. If student i's application score was lower than the cutoff score for his preference m, then he was rejected by major m. If his application score was higher than (or just at) the cutoff score for majors j, k, and l, and j was his preferred option, he was offered admission to major j, and rejected by majors k and l. With this information on hand, applicants then decide whether to enroll or not.

3.2.3 The Affirmative Action Policy

The Affirmative Action policy was implemented for the first time in 2007. The spirit of this policy was to give an edge in admissions to applicants who were top students in their high school and came from disadvantaged SES backgrounds. The rationale behind it relied in a corrective argument (Sandel 2009), rather than on an efficiency motive. Policy makers argued that low-income individuals that are top students at their respective lower-quality high schools are at an educational disadvantage when competing with higher SES students for spots at flagship universities. The implementation of the AA policy was therefore mainly perceived as a matter of fairness.[10]

The universities offered *extra* spots[11] to prospective students ranked in the top 5% of

---

9. As described in Kirkeboen, Leuven and Mogstad (2014), the truncation of the application list to a limited number of choices may threaten this statement if individuals use their last preference as a safe option. If that were the case, an important fraction of applicants should be admitted to their 8th choice. However, the data indicates that less than 0.6% of all applicants were admitted to the their 8th preference.

10. This policy originated in the context of large student movements in 2006. Under the pressure of the students' unions, the Ministry of Education and the flagship universities' Deans implemented the affirmative action policy I evaluate in this paper. Even though it represented a fairly moderate innovation, this was the first modification made to an otherwise unaltered university admission system since the 1960s. See Donoso (2013) for more details on the 2006 students' protests.

11. Like a similar AA policy implemented in Israel (see Alon and Malamud 2014), departments within

their high-school GPA who satisfied the following conditions: (i) graduated from publicly funded schools[12]; (ii) belonged to the first four national quintiles of household income; and (iii) had an application score below the cutoff score of their major of preference (on a 5% band). The income eligibility condition (ii) does not appear to be very restrictive, but it is coherent with the highly unequal income distribution in Chile,[13] mainly driven by its upper tail. In terms of household income, the annual average (in 2011 U.S. dollars) for the first quintile was about $4K; for the 2nd quintile, $8K; for the 3rd quintile $12K; for the 4th quintile, $18K; and notably for the 5th quintile, $51K. As I describe in the data section, AA beneficiaries come from households with significantly lower income than regularly admitted students.

The Ministry of Education handles all the information needed to determine the eligibility conditions (i) and (ii); they receive the data on high school GPA directly from high-schools, and the income data from the IRS. The borderline condition (iii) was imposed by the university officers, who did not want to admit seriously under prepared applicants. Therefore, they restricted eligibility to students that were only marginally rejected by the regular selection system. Suppose for example that the engineering cutoff score at university U was 700 points; then all those applicants who fulfilled the first two conditions and had an application score between 665.0 and 699.9 (the 5% band) could enroll using the AA policy. In practice, eligible applicants received the same admission results as the rest non-eligible students, consisting in their major-specific application scores with the respective cutoff scores. However, the AA eligible applicants also got an indication of which of their preferences offered them admission thanks to the policy. Therefore, the policy enlarged the choice set for eligible applicants. They could enroll either (i) using the AA policy, or (ii) through the regular admission, or

universities placed a cap on the number of entering students. As in Israel, I find that in the Chilean setup this restriction is not binding.

12. A 92% from all high-school graduates studied at publicly funded schools.

13. Only 8% of the population in the first four quintiles has a university degree, while 40% of the fifth quintile do (CASEN 2011).

they could opt out. Beneficiaries of the policy would not have been admitted to the major of preference had the policy not existed.

3.3 Empirical Strategy

3.3.1 *Institutions' Perspective: a 'Validity' Hypothesis*

The Chilean admission system is explicitly built on the premise that students with application scores above the major's cutoff would perform better than those below that threshold (DEMRE 2008). This rationale justifies that those below the cutoff are denied enrollment. Following Rothstein (2004) the assessment of student i in major j by the university officer may be written as

$$E[{Y_{ij}}^*|Z_{ij}] = \beta_0 + \beta_1 Z_{ij}$$

where ${Y_{ij}}^*$ represents a latent measure of academic success and Z_{ij} corresponds to the application score of individual i to major j. The problem confronted in the validity literature is that the performance of those who were not admitted is missing from the data, i.e., ${Y_{ij}}^*$ remains unobserved for students with an application score below the major j cutoff K_j. However, given the design of the AA policy I do observe ${Y_{ij}}^*|Z_{ij} < K_j$ for the set of AA beneficiary students.

Using equation (1) I compare mean differences in a host of academic outcomes - first year GPA, dropout rates and year-to-year persistence– between the AA beneficiaries and the group of regularly enrolled students that would have been displaced had the policy not consisted in extra slots. D_i equals one for beneficiaries and zero for the regularly enrolled students that would have been displaced by the beneficiaries. For illustrative purposes, imagine that six students enrolled using the AA policy in the major J. The potentially displaced students in this case correspond to the six individuals with the lowest application score that enrolled in major J, who I can identify from the individual data on applications and admissions. I will test $\gamma_m < 0$, i.e., whether beneficiaries perform worse than their higher-scoring peers on average.

$$\text{Mean Difference:} \qquad Y_{ij} = \beta_m + \gamma_m D_i + \epsilon_{ij} \tag{1}$$

$$\text{Difference at the Cutoff:} \qquad Y_{ij} = \beta_c + \gamma_c D_i + f(Z_{ij} - K_j) + \varepsilon_{ij} \tag{2}$$

This analysis is particularly policy relevant because it tells the university officers how the AA enrollees perform compared to the students that would have been denied access because of the policy. University officers were concerned about the quality of their majors because the policy was introducing students with lower application scores than regularly. This exercise is then a counterfactual analysis at the major level; without the policy, beneficiaries would have been rejected but with the policy they enrolled.

The admission score Z may be valid so long as some target outcome is declining (perhaps smoothly) as scores decline. I'm also going to test a version of equation (1) where $D_i = 1\left[Z_{ij} - K_j < 0\right]$ allowing for all regularly enrolled students to be part of the analysis. This exercise informs whether the admission score predicts performance and to what extent that prediction is different for beneficiaries and regularly admitted students.

To complete this analysis I will also estimate equation (2) where $f(Z_{ij} - K_j)$ is a smooth function of the application score, normalized so that the cutoff point equals zero for all majors . I use different distances to the cutoff and functional forms of $f(\cdot)$. According to the admission system rationale students that were just rejected (i.e. with scores just below the cutoff) should perform worse or at least equal to the ones above the cutoff. Hence, I am also interested in testing whether $\delta_c \leq 0$, which would support the admission system rationale. Finally, if $\delta_c > 0$ I would be estimating variations of equation (2) to find at what distance to the left of the admission cutoff beneficiaries perform worse or equal than regularly enrolled peers ($\delta_c \leq 0$). Given that the admission system relies exclusively on application scores, I do not include additional covariates in (1) or (2) to test its validity.

Additional Measures of Skill

At least to some extent, test scores do not measure skills such as motivation, persistence, and self-control (Heckman and Kautz 2012, 2014a). Previous research has shown that these skills are powerful predictors of relevant outcomes, including schooling attainment (Bowles, Gintis and Osborne 2001; Borghans et al. 2006; Heckman, Stixrud and Urzua 2006; Almlund et al. 2011, and Heckman and Kautz 2012). Hence, non-test score measures of skills may explain potential differences in academic outcomes between AA and regularly enrolled students.

Following Carneiro et al. (2003), Heckman and Kautz (2012) and Jackson (2014), I pose a factor model that extracts two underlying latent variables from a set of observed measures. I use the students' PSU scores in mathematics and language as a proxy for cognitive skill, and high-school GPA, high-school class rank, and high-school attendance as non-test score skills. These last variables have been used in the related literature (Heckman and Kautz 2012, Kautz and Zanoni 2015) as additional measures of skill, given that they reflect a behaviors that go beyond *pure* cognition, like showing up for class. Hence in this setup the cognitive skill factor[14] represents what is measured by the standardized tests and the non-test score skill factor is what is captured by the other measures. By relying on the covariance between measures described above, I identify the distribution of the two types of skills ($\widehat{\theta}_1$ and $\widehat{\theta}_2$), standardized to be mean zero unit variance.[15] I then augment the previous equations in the following way,

$$Y_{ij} = \beta_{ms} + \gamma_{ms} D_i + \gamma_1 \widehat{\theta}_1 + \gamma_2 \widehat{\theta}_2 + \mu_{ij} \tag{3}$$

$$Y_{ij} = \beta_{cs} + \gamma_{cs} D_i + f(\cdot) + \gamma_1 \widehat{\theta}_1 + \gamma_2 \widehat{\theta}_2 + u_{ij} \tag{4}$$

and test whether $H_0 : \gamma_{ms} < 0$ and $H_0 : \gamma_{cs} \leq 0$ after controlling for both types of skills θ_j

14. The assumption that one latent factor captures cognitive ability is traditional in the literature (see, e.g., Jensen 1998).

15. I use the Bartlett scoring method, but the results are robust to other methods. Note that I am not aiming to identify induced changes in skills overtime (as in Heckman et al. 2013), but establish pre-program 'baseline' differences in skills.

for j=1,2.

3.3.2 Students' Perspective: Access and Mismatch

Another important dimension of the policy is what would have happened to the AA eligible students had the policy not existed. This is a counterfactual analysis at the individual level. I ask two relevant questions on this respect. The first is how did the policy affect at which major students ended up, if any. The second is how eligible AA students perform in more selective majors compared to how they would have fared in less selective majors in absence of the policy, conditional on enrollment.

Access

I construct counterfactual enrollment options using individual preferences over majors from the application forms. Had the policy not existed, AA students could have either enrolled in the next preferred major that offered them admission under the regular system, or they could opt out.[16] It is also possible to have been rejected by every major listed on the application form.

This approach assumes that the behavior of the AA eligible students did not change as a response to the policy. This is reasonable to assume in a context in which the AA policy was announced (June) and implemented (December) almost immediately, leaving little space for students to react. I make and test the following two assumptions regarding the applicants' behavior.[17]

Assumption 1. Applicants did not anticipate the policy, and therefore did not switch to a different high school (e.g. with lower-achieving peers) to improve their chances of being in

16. Meaning, enrolling at technical or professional institutes, private less-selective universities, going to the labor market, or simply not enrolling at all.

17. I tested these two assumptions by analyzing the switching and applications behavior for different cohorts of students both before and after the policy was implemented. I examine the switching patterns and number of applications for students that were in the best 5%, 10% and 20% of their high school. I found no differences in either school switching or application behavior.

their class top 5%.[18]

Assumption 2. Students did not change their application behavior as a response to the policy.[19]

The availability of next-best alternatives allows to distinguish intensive and extensive margin effects as described in Arcidiacono and Lovenheim (2015). The intensive margin affects the group of applicants who would have been admitted to at least one alternative major and got reallocated thanks to the policy. Meanwhile, applicants who would not have been admitted to any major were affected at the extensive margin.

As described in the previous section, admissions are based on an algorithm that uses prospective students' application scores and their major preferences. Let student's i preference list be denoted by $P_i = \{j_{i1}, j_{i2}, ..., j_{i8}\}$, where j_{ik} represents a major, and k denotes the student's i preference order over majors. Students are required to list their preferences in order, so let $j_{i1} \succ j_{i2} \succ \cdots \succ j_{i8}$ where $\succ$ indicates strict preference relation. Without loss of generality, I assume that the preference relation can be represented by a utility value, so that $j_{im} \succ j_{in} \Leftrightarrow U(j_{im}) > U(j_{in})$. Let A_i denote the set of majors in which beneficiary i would have been admitted by the regular system. Had the AA policy not existed, student's i best option would have been given by ${j_{ik}}^*$, where

$${j_{ik}}^* = \text{argmax } U(j_{ik}) \quad \text{s.t. } j_{ik} \in A_i \subseteq P_i$$

Note that for some students I may have that ${j_{ik}}^* = \varnothing$, indicating that they would have been rejected by every other major they applied to. I use this piece of information to bound the extensive margin effect on enrollment as follows. Let C_i be a dummy variable defining

18. Cullen, Long and Reback (2013) showed that this assumption did not hold in the U.S. under Texas' top ten percent plan. In the Texas case, all students who graduated in the top 10% of their high school classes were *automatically* guaranteed admission to any in-state public higher education institution. In Chile, applicants need to comply with all the conditions stated in the preceding section, and even so, only students at the margin benefited.

19. This is reasonable given that the policy benefits 'marginal' students and the major's cutoffs are not known at the moment of applying. Moreover, evidence from Hastings, Neilson, Ramirez and Zimmerman (2015a) indicates that additional information has little impact on applicants' behavior in Chile.

student's i enrollment at flagship universities, and D_i be a dummy variable indicating if student i used the AA policy to enroll. As under a standard potential outcomes framework, C_{1i} defines student's i enrollment given $D_i = 1$, while C_{0i} defines student's i enrollment if $D_i = 0$. The impact on enrollment for the students using the AA policy is given by

$$E(C_{1i} - C_{0i}|D_i = 1) = P(C_{1i} = 1|D_i = 1) - P(C_{0i} = 1|D_i = 1)$$

in which $P(C_{1i} = 1|D_i = 1) = 1$ by definition, and $P(C_{0i} = 1|D_i = 1)$ can be bounded from the application data. I can decompose $P(C_{0i} = 1|D_i = 1)$ into two components, as

$$\begin{aligned} P(C_{0i} = 1|D_i = 1) = {} & P(C_{0i} = 1|D_i = 1, {j_{ik}}^* = \varnothing)P({j_{ik}}^* = \varnothing|D_i = 1) \\ & + P(C_{0i} = 1|D_i = 1, {j_{ik}}^* \neq \varnothing)P({j_{ik}}^* \neq \varnothing|D_i = 1) \end{aligned}$$

where $P({j_{ik}}^* = \varnothing|D_i = 1)$ and $P({j_{ik}}^* \neq \varnothing|D_i = 1)$ are observed probabilities. Students who would have been rejected by every major they listed on their application form would not have been able to enroll under any circumstance in flagship universities, so $P(C_{0i} = 1|D_i = 1, {j_{ik}}^* = \varnothing) = 0$. The group of students who would have been admitted to alternative majors could either have enrolled at their counterfactual choice or not, hence $P(C_{0i} = 1|D_i = 1, {j_{ik}}^* \neq \varnothing) \in [0, 1]$. Therefore,

$$P(C_{0i} = 1|D_i = 1) \in [0, P({j_{ik}}^* \neq \varnothing|D_i = 1)]$$

and hence I can bound the impact of the policy on pure enrollment as:

$$E(C_{1i} - C_{0i}|D_i = 1) \in [1 - P({j_{ik}}^* \neq \varnothing|D_i = 1), 1] \tag{5}$$

The lower bound is zero if all beneficiaries are admitted to alternative majors under the regular admission system. The lower bound is higher than zero (and lower than one) if the policy allows some students who would have been otherwise left-out to enroll at flagship universities.

For the group of beneficiaries who would have been admitted to alternative majors I can compare further outcome dimensions. For instance, I can answer whether the policy re-sorted students to majors with higher expected wages, or to majors offered by more prestigious institutions.[20] These two measures of university-major quality provide us an aggregated notion for understanding how expanded access benefited AA students.

Student Perspective: a Mismatch Hypothesis

The *mismatch hypothesis* (Sander 2004, and Sander and Taylor 2012) states that AA eligible students would have fared better in other –less selective– majors. The strategy to assess the mismatch hypothesis in this context relies on the discontinuity in the eligibility for the AA policy generated by the class rank condition. I consider students that fulfilled all the eligibility conditions with the exception of class rank, and compare those just above and just below the top 5% rank cutoff. Therefore, the intention-to-treat identifies how would the marginally rejected low-income students perform in their majors given that they were class-rank eligible.[21] The identification assumption is that the group ranked just below the cutoff provide a good counterfactual of students ranked above. I expect students near the 5% threshold to be similar on their observed and unobserved characteristics, but have very different likelihoods of becoming beneficiaries of the AA policy.

I employ a regression discontinuity (RD) design to compare outcomes between students

20. To this end, I use a single summary measure of university prestige perception, similar to the Barron's Index for U.S. institutions. It is based on a survey given to Chilean CEOs and managers from different departments, areas and firms. Basically, they were asked to evaluate universities in a scale ranging from 1 to 7, in which a higher value means greater prestige.

21. Admittedly, I could have used a different comparison sample. For instance, I could have used the income variable to perform an RDD using the discontinuity between the fourth and fifth quintiles, but the information on income was provided to us only grouped in quintiles. I could also have included all students, not only those that met the application score band condition. I decided to keep this requirement because of interpretability. This way, the ITT provides us the effect for low-income students that would have been eligible for the policy if they belonged to the best 5%. Otherwise, it would be conflating the effect for those that met and those that did not meet the confidence band requirement.

who attend selective majors with their counterparts who attend less selective ones near the class rank cutoff. The specification has the following form,

$$Y_{ij} = \alpha + \delta \cdot Best5_i + f(rank_i) + X_i'\beta + \varepsilon_{ij} \tag{6}$$

in which Y_{ij} represents an academic outcome (like persistence rates) for student i at major j, $Best5_i$ is an indicator function for student i ranked in the top 5% of their high school, $f(rank_i)$ is a smooth function of the forcing variable class rank and X_i represents a set of observed, arguably exogenous, covariates. These include parental schooling, household income, health system use (public or private) gender, GPA and PSU scores and high school quality, which I describe in the data section. In practice, these controls do not change the estimate at the discontinuity but increase its precision. The parameter δ is the coefficient of interest. I expect it to be negative, given that eligible students enroll at more selective majors which tend to be more challenging in the Chilean context. If the coefficient is also significant, then we would be in the presence of 'mismatched' students, who would have fared better in less selective majors.

I specify a flexible parametric model by including different higher order polynomials of the forcing variable, which are allowed to vary on either side of the discontinuity. I also use non-parametric specifications, following Hahn et al. (2001) and Porter (2003), and estimate a local (kernel-weighted) linear regression to the left and right limits of the discontinuity. As suggested by Lee and Lemieux (2010), I examine the results for a range of candidate bandwidths, including an optimal Cross-Validation (Ludwig 2007) and the Imbens-Kalyanaraman bandwidths. Finally, all regressions cluster at the level of high school in order to avoid the problems associated with specification error in the case of discrete covariates (Lee and Card 2008). Another key assumption for the RD design is that students were not able to manipulate their class rank around the eligibility threshold. This non-manipulation is related to the assumption 1 from the previous section. In this setup, I test for manipulation by analyzing systematic patterns in the density on either side of the class rank threshold, using McCrary's (2008) procedure.

3.4 Data

I collected information from several data sources to implement my empirical strategy. I gathered administrative data sets with individual records on high-school GPA and attendance, and high-school characteristics. I combined this information with the students' PSU scores and application lists to flagship universities. Finally, I also gathered administrative records on student university enrollment, up to five years after high-school graduation.

I use the administrative data on university enrollment to track students yearly after initial enrollment and construct several measures of academic success using this information. The first is a strict measure of persistence, defined to equal one for students that stay in their same pair university-major from year to year and zero otherwise. This is an important outcome for university officers, who are concerned about admitting students that would finally graduate from their programs. I also got degree completion records for students in the 4th and 5th year, which I use to avoid conflation between dropout and graduation in the construction of this persistence variable. However, I do not use graduation itself as an outcome because the formal length of the majors in Chile is on average above 5 years and therefore only a small fraction of students had graduated by the 5th year. For computation procedures, I count the students that graduated early (i.e. in years 4 or 5) as persisted.

I also build measures of continuous enrollment, stop-out and drop-out rates, in the spirit of Stratton et al. (2008). Previous evidence (Pugatch 2012 for South Africa, Arcidiacono et al. 2013 and Goldrick-Rab 2006 for the U.S.) shows that a substantial fraction of students withdraw during the first year and then return to postsecondary institutions, which is a pattern also found recently in Chile (Mineduc 2014b). Considering the five years of data, I define continuous enrollment as enrollment for 4-5 years; students in the category of stop-out as those who were enrolled for 2 or 3 years, and dropout for those enrolled only 1 year or less.

Finally, for a subsample of students I was able to collect 1st year GPA for a subset of institutions. This information is usable only for the validity analysis because I would need to have information on 1st year GPA for students in different pairs majors-institutions to perform the mismatch analysis. Table 3.16 in the Appendix shows that both AA beneficiaries and Regularly enrolled students in this subsample are fairly similar to the complete sample described below.

The final dataset contains observations from the 2007 high-school graduating cohort who took the PSU exam, applied to, were admitted by, and enrolled as freshmen at flagship universities. I can identify all AA eligible students, those that become beneficiaries, and all regularly enrolled students, with complete information determining their admission to the respective university-major pair. A total of 137K students graduated from publicly funded schools and took the PSU test; 41K applied to flagship universities, and approximately half of them (21K) satisfied the income criteria. About 12% of them ($\sim$2.5K) satisfied the application score band requirement for at least one of their listed preferences. These 2,478 applicants compose the sample I use to study the mismatch hypothesis. Among them, 713 of them were in the top 5% of their high school class, thus eligible for the AA policy and 466 (two out of three) became beneficiaries.

A total of 18,790 students enrolled under the regular system to the same majors as beneficiaries. From this group, a subset of 475 students that would have been displaced by the 466 beneficiaries. The slightly higher number of potentially displaced students is due to individuals with the same the application score, who I also kept in the sample. I use beneficiaries and regularly enrolled students as the sample to test validity hypothesis associated to the admission system. Table 3.1 presents descriptive statistics for the validity and the mismatch samples.

Validity Sample. Next, I describe the validity sample by comparing all of the regularly enrolled in the same majors[22] (in column [1]), the students that would have been displaced (in column [2]) and the AA beneficiaries (column [3]) .

The set of available indicators reflects significant differences in socioeconomic status between beneficiaries and the regularly enrolled students, accordingly with the targeted purpose of the policy design. AA beneficiaries have parents with lower levels of education (12.6 years vs. about 14 years of schooling[23]), and come from households with significantly less household income. Three-quarters (74%) of the AA beneficiaries use the public health system, which is related to low SES status[24] (Larranaga and Contreras, 2010) compared to about half (49%) of regularly enrolled students.

All AA students graduated from publicly funded schools, compared to 73% of their classmates. The quality of high-schools from which AA students come from is also lower. They come from average high-schools, while regularly enrolled students graduated from schools averaging a score of 298, which equals the 80th percentile of the high-school SIMCE distribution. Also, AA students present significantly lower PSU scores. They achieved 598 points, which is 50 points below the mean of the regularly enrolled group (about 0.8 standard deviation of difference).

The information coming from the application forms indicates that AA beneficiaries applied on average to 5.6 university-major combinations, while regularly enrolled applied to 4.9. Thanks to the AA policy, beneficiaries were allowed to enroll approximately at their second preference (1.9), on average.[25] As a benchmark, students that were regularly enrolled to the same majors ($N = 18,790$) listed 4.9 preferences, and were admitted to their second

22. The AA students enrolled in a total of 286 university-major combinations. The small sample size per combination prevents us from performing a subgroup analyses by university-major categories.

23. As a benchmark, 12 years of schooling is equivalent to have completed secondary education in Chile.

24. The national health care system in Chile is highly stratified by its public/private type. Using the public health system is much cheaper than the private sector, which is run by for-profit health insurance companies.

25. This preference of enrollment for AA beneficiaries considers 392 observations instead of 466, because 74 of them would not have been not admitted to any major under the regular admission system, as I elaborate in section 3.5.

preference ($\sim$1.8), on average.

According to their SES and high-school quality indicators and PSU scores AA students are expected to perform worse than regularly enrolled students at the university. However, this is not as clear-cut in terms of the rest of their academic credentials. AA beneficiaries are by design, top ranked students from their high schools. On average, their class rank is 0.98, significantly higher than the 0.79 of the regularly enrolled students. They also have higher high school GPA (713 vs. 667 points, or about 0.6 standard deviation of difference). Still, in the literature of economics of education it is common to find that SES indicators are by far the most important determinants of academic achievement. It is not at all obvious that good GPA or class rank indicators would allow AA beneficiaries to compensate their SES background disadvantages.

The panel at the bottom of Table 3.1 shows the fraction of students from each group in enrollment categories (continuous, stopout and dropout), and the year-to-year persistence rates through freshmen to the fifth year at university.

The information from the enrollment categories indicates that 82% of the AA beneficiaries stays enrolled for 4 or 5 years, lower than the 85% of the regularly enrolled students, but similar to the potentially displaced students. The stopout and dropout rates are fairly similar between these last two groups indicating that AA beneficiaries manage to achieve an academic performance that is at least similar to the ones they would have displaced.

This notion is strongly reinforced by their yearly persistence rates. The numbers show that 83% of the AA beneficiaries stay in the same major-institution by the end of the first year, which is equivalent for the regularly enrolled students and higher than what the potentially displaced students achieve (80%). The table indicates that AA beneficiaries tend to have higher persistence rates as they advance in their degree. By the fifth year, 60% of AA beneficiaries stay enrolled (or graduated) from the major they initially enrolled on, compared to 57% and 54% of the other two groups. This result suggests that the indicators of higher ability (class rank and higher GPA), offset the lower PSU scores, high-school quality and socioeconomic disadvantage faced by AA beneficiaries.

Mismatch Sample. The characteristics of students in the mismatch sample are described in columns [4] and [5] from Table 3.1, separately according to non-eligible and eligible status, respectively. The difference between columns [4] and [5] is the class rank condition; those in the second column correspond to those who ran in the top 5% of their class, i.e. were eligible for the AA policy. Accordingly, the average class rank for the non-eligible is of 0.63, while the mean is 0.98 for those eligible. This academic advantage in favor of the eligible students is also reflected in their higher high-school GPA and PSU scores. Given that all individuals in the mismatch sample satisfied the income condition, differences between eligible and non-eligible students are less evident when looking at socioeconomic indicators. They tend to come from households with similar levels of parental education and income. Even though all students in the mismatch sample graduated from publicly funded high-schools, eligible students come on average from lower quality high-schools, as measured by Chile's national standardized testing system, SIMCE.[26] The fact that class rank correlates negatively with SES status and high-school quality arises naturally given that just the very good students from poor schools make it to the mismatch sample. Finally, eligible students tend to be women living in the capital region in a higher proportion, and more likely to get a scholarship. In spite of the differences, all these pre-university variables are smoothly distributed across the class rank threshold, as I present and describe in Section 3.5.3. In terms of applications, both eligible and non eligible applied to an average of about six alternatives, and enrolled in their third (non-eligible) or second (eligible) preferred option.

Finally, students who are AA eligible tend to be more time continuously enrolled and stopout less. They also have systematically higher persistence rates than non-eligible students, and that the gap increases over time. This is not a surprising result as AA eligible

26. SIMCE stands for Education Quality Measurement System (Sistema de Medicion de la Calidad de la Educacion in Spanish). Students take mandatory tests in Mathematics and Language, that then are averaged at the school level to produce the school's SIMCE score. I use the latest SIMCE measure for high-schools for that date, corresponding to years 2006-2008. The average SIMCE score for schools was 257 with a standard deviation of 40 points.

students have both academic and SES indicators that are higher than non-eligible students. For the mismatch sample, the relevant result is whether AA eligible students perform better or worse than other highly ranked (say, those in the top 6%) but non-eligible students, as I examine in the RDD results subsection.

3.5 Results

3.5.1 Validity

Table 3.2 reports mean differences in a host of measures of academic performance between the AA beneficiaries and the regularly enrolled students that would have been displaced by them. Panel A exhibits the results from separate estimations as specified in equation (1), using 1st year GPA (for a subsample) and year-to-year persistence rates as outcomes.

These outcomes are the most relevant from the perspective of the validity analysis, since university officers aim in first place to admit students that will achieve a good performance in their own programs.

The table reports both the standard errors (in brackets) and the lower one-sided p-values (in parentheses) associated to the coefficients on the beneficiary status dummy. The results indicate that AA beneficiaries do not perform worse on either 1st year GPA or yearly persistence rates. In fact, the coefficients are all positive and also significant (though at the 10% level) for the fifth year, suggesting that AA beneficiaries tend to graduate early in higher proportion.

Panel B presents results from a simple multinomial logit model, using enrollment behavior as the outcome and the beneficiary dummy as independent variable. This outcome is less strict than remaining in the same pair major-institution yearly but it may shed light on a broader measure of persistence. Columns (1) and (2) report odds ratios and standard errors comparing dropout and stopout with continuous enrollment, while column (3) shows the odds ratio and standard error associated to the comparison of dropout with the stopout category as base outcome. The results indicate that there are no differences between both groups, according to the Wald Test with a chi-squared of 0.236, which is consistent with the sample means on enrollment behavior for both groups presented in Table 3.1.

This set of results questions the validity of the admissions score since students that would have been rejected attain good indicators of performance. Figure 3.1 provides further evidence to understand how the admission index predicts performance for both groups below and above the admission cutoff. The set of graphs plot the mean of yearly persistence rates within admission score bins, and the solid line plots fitted values from local linear regressions on both sides of the admissions cutoff.

The graphs suggest that persistence rates tend to be flat in the application score for regularly admitted students, and that AA students perform better near the thresholds. These results are confirmed by the the output presented in Table 3.3. The table reports the estimated γ_c from separate regressions of persistence rates on a beneficiary status dummy and a smooth function of the normalized application score as described in equation (2). The first panel continues to use the sample of potentially displaced students to compare to the AA beneficiaries. Panels B and C include other regularly admitted students near the admission cutoff to estimate the difference at the thresholds using non-parametric and parametric procedures, respectively.

The results, which are stable across the different specifications and procedures, indicate AA beneficiaries achieve significant higher persistence rates than their regularly enrolled peers at the cutoff. The graphs from Figure 3.1 suggest that AA beneficiaries achieve higher persistence rates that up to 8 points below the admission threshold.

The fact that the AA indicator does predict persistence across a range of test scores despite not being included in the admissions index suggest that if the goal of the assignment process is to maximize persistence, institutions could improve the indices they use to determine admissions outcomes by incorporating class rank.

While the validity premise relies solely on the application score, it is illustrative to examine how beneficiaries differ from regularly enrolled students on their demographic, socioeconomic and pre-university academic characteristics near the admission cutoff. Table 3.4 reports the results from separate regressions of each covariate on a beneficiary status dummy and a smooth function of the normalized application score, once again separated in Panels A, B and C. Results from the three panels display the following consistent pattern. While there are no differences in gender close to the admission cutoffs, the mean differences on the rest of the covariates persist near the admission cutoffs. Even close to the admission cutoffs, AA students attended high schools of lower quality; have lower levels of parental education, come from households with lower income, have higher high-school GPA, use the public health system in higher proportion, are less likely to live in the capital region, and are more likely to have scholarships to fund their studies. Figures 3.2 and 3.3 provide a visual illustration of these differences.

Additional Measures of Skill. Table 3.5 presents a correlation matrix of the pre-university skill measures used in the factor model posed in Section 3.3. It displays patterns that are common to other contexts (e.g. Jackson 2014, Kautz and Zanoni 2015), namely, that test scores correlate positively with each other (.46), and that test scores have a a weaker correlation with other measures of skill, like high-school GPA (about .26), high-school class rank (.11) and high-school attendance. In this last case the correlation is close to zero (.01 for math and -.05 for language). Even though unrelated to test scores, showing up for class is related to higher GPA (.13) and higher class rank (.18) in high school. These last two measures exhibit a high correlation (.70), but still different from 1. This result indicates that there are students from different high-schools with same GPA but different relative position within their graduating class. Table 3.6 shows the correlations between the aforementioned measures of skills and persistence rates at the university. GPA and class rank show the highest correlation with persistence rates, which tends to increase as students progress in their majors. This is in contrast with PSU scores, whose correlation with persistence rates tend to stay constant or decrease over time. High-school attendance correlates with persistence rates at the university despite its petty relation with test scores. In fact, the magnitude of the attendance correlations with persistence rival with the test scores figures. Taken together, this suggest that test scores and attendance measure different skills, both of which are related to future outcomes. All in all, Tables 3.5 and 3.6 indicate that measures other than just test scores capture information on skills that seem important in explaining academic persistence. Figure 3.4 presents the distributions of predicted skill factor scores associated to test-scores (left panel) and non-test-scores (right panel), by AA status. The former are based on the subscores from the mathematics and language PSU test, while the non-test-scores skill factor is based on high-school GPA, class rank and attendance rates. The graphs show that AA beneficiaries possess lower cognitive skills, but higher non-test score skills. Tables 3.7 and 3.8 show the results of the estimation of equations (3) and (4), in which I control for both the test-scores ($\widehat{\theta}_1$) and non-test-scores skill ($\widehat{\theta}_2$) factors. The results show that once controlling for these skills the advantage favoring AA students vanishes for most of the specifications. This fact suggests that the low-income AA students are equipped with a set of skills –not adequately measured by the admission system– that enable them to do well in post-secondary education.

3.5.2 Students Perspective: Access

The next subsection moves beyond from the university officers perspective to examine the policy from the perspective of the individual students. Instead of comparing AA beneficiaries to their peers, I compare now AA-eligible students to their own counterfactual in the absence of the policy. Table 3.9 summarizes how the policy expanded the enrollment opportunities for the 713 AA eligible students. The rows indicate the order of preference from student's i best next-best alternative (j^*_{ik}), while the columns indicate the order of preference at which they ended up enrolling. The 247 eligible students who did not become beneficiaries appear on the diagonal, which indicates that they did not need the AA policy to enroll at their best option. Meanwhile, the 466 students under the diagonal correspond to the AA beneficiaries, who were able to enroll at a more preferred university-major combination (or enroll at all, as I explain below). The second row-first column cell shows that 159 AA students were admitted to their second preference under the regular system, but used the policy to get enrolled in their first preference. Next, 91 AA students were admitted to their third preference (and not to their first nor second preference). Of these, 45 were able to enroll in their first preference thanks to the policy and 46 were able to enroll in their second preference. The following rows can be read in the same way.

The row for which $j^*_{ik} = \varnothing$ provides important information. It indicates that 74 AA beneficiaries would not have been admitted to any of their declared application preferences under the regular system. However, thanks to the policy, 21 of them were able to enroll in their first preference, 17 their second preference, and so on. Therefore, 392 AA beneficiaries would have been admitted to alternative majors, while 74 of them have been rejected by every major on their preference list. Hence, $P({j_{ik}}^* = \varnothing | D_i = 1) = 74/466 = 0.16$ and therefore the pure enrollment (or extensive margin) effect as described in expression (1) can be bounded as $\overline{C}_{1,D=1} - \overline{C}_{1,D=0} \in [0.16, 1]$. This result means that at least 16% of the beneficiaries would not have enrolled at a flagship university had the policy not existed. This extensive margin effect could scale to a 100% if those who would have been admitted to alternative majors still decided to opt out.

For the group of potentially reallocated students -the intensive margin of the policy- I can compare differences in quality between their factual and their counterfactual major (j^*_{ik}). Table 3.10 provides examples. Consider the first case (first row). A particular student used the AA policy to enroll in Law at Universidad de Chile for which he presented an application score 10 points below the major's cutoff. Had the policy not existed, his next-best alternative would have been Law at Universidad Catolica de Valparaiso, since for that major he had a score 49 points above the cutoff. I compare differences between the factual and conterfactual majors using expected wages (at the major level) and a prestige index (at the institution level) as measures of quality.[27] Table 3.11 shows the differences in these two measures of major quality. The results indicate that beneficiaries were allowed to enroll in majors with higher wage expectations and at more prestigious universities.

3.5.3 Mismatch Sample

This subsection analyzes how AA-eligible students perform compared with themselves in the absence of the policy, using an RDD on the class rank variable. Since the academic outcomes of students are available for those who enroll, I discard students at the extensive margin from this analysis. This might moderate the potential selection bias introduced by keeping in the sample students that would not have been able to enroll in the absence of the policy (i.e., those benefited at the extensive margin). Using this sample, Figure 3.5 plots the proportion of students who were beneficiaries of the AA policy against their class rank, revealing a discrete jump in the probability of becoming a beneficiary near the class rank threshold.

The main identifying assumption behind this strategy is that there are no other changes occurring at the top 5% class rank threshold that could confound the analysis. I test for these differences by regressing a host of different covariates on the class rank dummy identifying top 5% students and a smooth function of class rank. As indicated in section 3.3.2, I consider both parametric and non-parametric functions of class rank to ensure that the results are robust to specification. The results of these regressions are presented in Table 3.12.

The reported coefficients are for the variable 'Best 5%', defined as 1 for individuals in the top 5% from their high school. The coefficients indicate that gender, high school quality, parental education, household income, high school GPA and PSU scores, public

27. The first measure corresponds to the average wages associated to the respective major for individuals in the fourth year after graduation. The data is collected from IRS tax records, processed by the Ministry of Education and published online at the institutional webpage *http://www.mifuturo.cl*. Regarding the prestige index, I use a university ranking published yearly by a Chilean weekly news magazine that is similar to the Barron's index for U.S. institutions. The index is based on survey information provided by business managers, who make hiring decisions at their firms.

health use, among others, evolve smoothly through the best 5% threshold. This result holds for different local-linear bandwidths and polynomial functions of class rank.[28] The set of graphs in Figures 3.7 and 3.8 provide a visual check of these results by showing how the same set of covariates behave with class rank. The dots plot the mean of each variable within class rank bins, and the solid line plots fitted values from local linear regressions on both sides of the class rank cutoff. The graphs confirm that demographic, socioeconomic and pre-university academic variables behave smoothly around the class rank threshold, which suggest that unobserved characteristics may display a similar behavior.

One way to further test for differences in unobserved characteristics is to examine whether there is sorting of the sample around the discontinuity. For example, it could have been the case that more motivated students switched to easier high-schools in order to be eligible. I employ McCrary's (2008) density test to examine the existence of sorting in the vicinity of the class rank threshold. Figure 3.6 presents a graphical illustration of this test, which shows that the density varies continuously over different class rank values with no significant discontinuity around the threshold.[29] Overall, the evidence provided above indicate that the students' characteristics are not related to the changes near the class rank threshold other than the likelihood of becoming beneficiary of the AA policy.

I examine yearly persistence rates in the same pair major-institution and also dummies reflecting whether students were continuously enrolled (independent of where) or if they exhibited a dropout behavior, meaning that they were enrolled for one or less years during a five years period. I do not include as outcome a dichotomous variable on stopout behavior since it is more difficult to interpret.

---

28. In Table 3.12 the panel A presents the non-parametric estimations, with local-linear bandwidths of 1, 3 and 5 percentiles of class rank (arbitrarily selected), and optimal bandwidths computed according to the Cross-Validation (CV) (Ludwig and Miller, 2007) and the Imbens-Kalyanaraman (IK) methods, respectively. Panel B shows the parametric estimations corresponding to linear, quadratic and cubic polynomial functions for $f(rank_i)$, allowing slopes to vary at either sides of the class rank cutoff. The results indicate that among the 88 estimated specifications only two reported significant coefficients at the 5% level, corresponding to gender for the linear spline and PSU score for the cubic spline.

29. McCrary's test indicates that the discontinuity estimate is given by .00323, with associated standard errors of .00270 (and hence non-significant).

Table 3.13 shows the differences in persistence rates from freshmen to the fifth year at university near the threshold. It reports the estimated δ from the specification (2), with the yearly persistence rates as outcomes in each column and separated panels for non-parametric and parametric procedures. The coefficients are negative which is consistent with eligible students attending more difficult majors. However, the vast majority of the coefficients are not significant at conventional levels.[30] For illustrative purposes, Figure 3.9 plots the persistence rates achieved by students as they progress through university, from freshman to senior year and their graduation rates.[31]

Table 3.14 reports the results on dropout behavior and continuous enrollment. As in the previous set of results estimates are imprecise, but in this case, almost all of the coefficients suggest that AA-eligible students dropout less and stay continuously enrolled in a higher proportion.

An hypothesis that reconciles both set of results is that on a yearly basis, AA students perform better in less selective majors due to the relative challenge that is pursuing a more selective major. However, considering a long term measure of academic success they AA students tend to do better in more selective majors because they are persistent and catch up with their peers over time, as results from the validity section showed.

30. The exceptions are the persistence rates at second year from the non-parametric estimations with 3 points of bandwidth, and the parametric cubic spline, both significant at the 5% level. I am aware that RDD is an empirical approach with low statistical power. All regressions control for the set of covariates displayed in Table 3.12, which increases precision of my estimates without affecting their magnitudes. I use different bandwidths to test whether effects become significant as I include more observations (trading variance for bias), which is not the case.

31. It might be the case that persistence rates vary importantly between majors. Therefore, I also consider relative persistence rates as outcomes. I measure relative persistence as the distance in persistence as compared to the mean from their respective major. The results, shown in Table 3.15 suggest that in relative terms, AA students perform somewhat better than in absolute terms. Still, as in Table 3.13, most coefficients are not statistically significant.

3.6 Conclusion

This paper evaluated whether test-score-based admission standards screen out highly able but low-socioeconomic status applicants. I used data from a Chilean Affirmative Action policy to measure outcomes on university access, mismatch and validity. The Chilean centralized, score-based university admissions system provides us the complete information used in determining the applicants' admission to their respective majors and universities and their next-best alternatives, which I combine with rich institutional administrative records. The results indicate that about 500 hundred students (two thirds of the eligible set) enrolled at using the AA policy. From them, 84% were reallocated to majors with higher wage expectations, at more prestigious universities due to the policy. Meanwhile, 16% of the beneficiaries would not have been admitted to any major at all, had the policy not existed.

Programs that promote access would not have any significant effect on educational attainment if they attract students who are unable to perform well or graduate. The mismatch hypothesis argues that beneficiaries may be so misplaced academically that they would actually be better off in the absence of affirmative action (Sander, 2004). I provide evidence that is not the case in Chile. Using a regression discontinuity design on the class rank condition, I do not find strong evidence supporting the mismatch hypothesis. The paper also contributes to the 'validity' literature (Rothstein, 2004). I take advantage of the policy design to test the admission system logic, by comparing beneficiaries' academic outcomes with regularly enrolled students in their same majors. The admission system predictions indicate that AA students would perform worse than regularly enrolled students, due to their lower application test scores. However, the data rejects this statement. The results indicate that on average AA students perform at least as well as the pool of regularly enrolled students in the same majors. Moreover, when comparing students closer to the admissions cutoff, I find that AA students attained sizable higher persistence rates compared to regularly enrolled students. The paper finally contributes to a growing literature relating non-test-score skills and educational attainment. My findings reveal that AA students have a set of skills that allow them to compensate for their low SES status and their initial knowledge handicap, reflected in their low PSU scores and in the lower quality of the high-schools they graduated from.

Summarizing, the AA policy contributed to equity by expanding opportunities for low income-students, otherwise rejected by the school-major combination they ended up attending. In the absence of the policy, they would not have been better off, and moreover, they exhibit a good academic performance relative to their peers. The set of results suggests that the admission systems could be improved by considering additional measures of skills. This paper contributes to the lively policy debate on affirmative action, suggesting that there is an opportunity to move forward on both equity and efficiency in university admissions.

3.7 References

Abdulkadiroglu, A., J. Angrist, and P. Pathak, 2014. The Elite Illusion: Achievement Effects At Boston And New York Exam Schools. Econometrica, Vol. 82, No. 1 137-196.

Almlund, M., Duckworth, A., Heckman, J.J., Kautz, T., 2011. Personality psychology and economics. In: Hanushek, E.A., Machin, S., Wossmann, L. (Eds.), Handbook of the Economics of Education, vol. 4. Elsevier, Amsterdam, pp. 1-181.

Alon, Sigal and M. Tienda, 2005. Assessing the 'Mismatch' Hypothesis: Differences in College Graduation Rates by Institutional Selectivity, Sociology of Education, 78, pp. 294-315.

Alon, S. and O. Malamud, 2014. The impact of Israel's class-based affirmative action policy on admission and academic outcomes. Economics of Education Review, Volume 40, pp 123-139.

Antonovics, K. and B. Backes, 2014. The effect of banning affirmative action on college admissions policies and student quality. Journal of Human Resources.

Arcidiacono, Peter and Michael Lovenheim, 2015. Affirmative Action and the Quality-Fit Tradeoff. NBER Working Paper 20962.

Arcidiacono, Peter, Esteban Aucejo, Patrick Coate, and V. Joseph Hotz, 2014. Affirmative Action and University Fit: Evidence from Proposition 209. IZA Journal of Labor Economics, 3:7..

Backes, B., 2012. Do affirmative action bans lower minority college enrollment and attainment? Evidence from statewide bans. Journal of Human Resources, 47(2), 435-455.

Belasco, Andrew S., Kelly O. Rosinger, and James C. Hearn, 2014. The Test-Optional Movement at America's Selective Liberal Arts Colleges A Boon for Equity or Something Else?. Educational Evaluation and Policy Analysis.

Blume, G. and M. Long, 2014. Changes in Levels of Affirmative Action in College Admissions in Response to Statewide Bans and Judicial Rulings. Educational Evaluation and Policy Analysis 36(2), 228-252.

Bordon, P. and Chao Fu, 2015. College-Major Choice to College-Then-Major Choice, Forthcoming at the Review of Economic Studies.

Borghans, L., Duckworth, A. L., Heckman, J. J., and Ter Weel, B., 2008. The economics and psychology of personality traits. Journal of Human Resources, 43(4), 972-1059.

Bowles, S., Gintis, H., and Osborne, M., 2001. The determinants of earnings: A behavioral approach. Journal of Economic Literature, 1137-1176.

Carneiro, Pedro, Karsten Hansen, and James J. Heckman. 2003. Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. International Economic Review 44, no. 2:361-422.

Cashin, S., 2014. Place, not Race: A New Vision of Opportunity in America. Beacon Press.

Contreras, D., S. Gallegos and F. Meneses, 2009. Determining Factors Of University Performance: Is Relative Ability Important? (Determinantes De Desempeno Universitario: Importa La Habilidad Relativa?) Revista Calidad En La Educacion, N30.

Cullen, Julie, Mark C. Long, and Randall Reback, 2013. Jockeying for position: Strategic high school choice under Texas' top ten percent plan. Journal of Public Economics 97 (2013) 32-48.

Daugherty Lindsay, Paco Martorell and Isaac McFarlin Jr., 2014. Percent plans, automatic admissions, and college outcomes. IZA Journal of Labor Economics, 3:10.

DEMRE, 2008. Admission Rules. DEMRE Series, Chile.

Donoso, Sofia, 2013. Dynamics of Change in Chile: Explaining the Emergence of the 2006 Pinguino Movement. Journal of Latin American Studies, 45, pp 1-29.

ETS (Educational Testing Service), 2005. Evaluacion Externa de la Prueba de Seleccion Universitaria.

Fischer R. and Repetto A., 2003. Selection Method And Academic Results. Engineering Department Of Universidad De Chile. Cep. Public Sector Studies N 92.

Fletcher, Jason M. and Adalbert Mayer, 2014. Tracing The Effects Of Guaranteed Admission Through The College Process: Evidence From A Policy Discontinuity In The Texas 10% Plan. Contemporary Economic Policy, Western Economic Association International, vol. 32(1), pages 169-186, 01.

Hahn, J., Todd, P., and Van der Klaauw, W., 2001. Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica, 69(1), 201-209.

Hastings, Justine S., Christopher A. Neilson and Seth D. Zimmerman, 2013. Are Some Degrees Worth More than Others? Evidence from college admission cutoffs in Chile. NBER Working Paper 19241.

Hastings, Justine S., Christopher A. Neilson, Anely Ramirez and Seth D. Zimmerman, 2015a. (Un)Informed College Choice: Evidence from Chile. Forthcoming. Economics of Education Review.

Hastings, Justine S., Christopher A. Neilson and Seth D. Zimmerman, 2015b. The Effects of Earnings Disclosure on College Enrollment Decisions. NBER Working Paper 21300.

Heckman, James J., Jora Stixrud and Sergio Urzua, 2006. The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. Journal of Labor Economics, Vol. 24, No. 3.

Heckman, J. J. and T. Kautz, 2012. Hard evidence on soft skills. Labour Economics 19(4), 451-464.

Heckman, J. J. and T. Kautz, 2014a. Achievement Tests and the Role of Character in American Life, In The Myth of Achievement Tests: The GED and the Role of Character in American Life, edited by J.J. Heckman, J.E. Humphries, and T. Kautz. Chicago, IL: University of Chicago Press. pp. 3-56.

Heckman, J. J. and T. Kautz, 2014b. Fostering and measuring skills: Interventions that improve character and cognition. In J. J. Heckman, J. E. Humphries, and T. Kautz (Eds.), The Myth of Achievement Tests: The GED and the Role of Character in American Life, pp. 34-430. Chicago, IL: University of Chicago Press.

Hinrichs, P., 2012. The effects of affirmative action bans on college enrollment, educational attainment, and the demographic composition of universities. Review of Economics and Statistics, 94(3), 712-722.

Hoxby, Caroline and Christopher Avery, 2013. The Missing One-Offs: The Hidden Supply of High-Achieving, Low-Income Students. The Brookings Institution, vol. 46(1 (Spring), 1-65.

Jackson, C. Kirabo, 2014. Non-Cognitive Ability, Test Scores, and Teacher Quality: Evidence from 9th Grade Teachers in North Carolina. NBER Working Paper 18624. Revised October 2014.

Kahlenberg, Richard D., 2014. The Future of Affirmative Action: New Paths to Higher Education Diversity after Fisher v. University of Texas. The Century Foundation Press, June 17, 2014.

Kapor, Adam, 2015. Distributional Effects of Race-Blind Affirmative Action. Job Market Paper, Yale University.

Kautz, Tim and Wladimir Zanoni, 2015. Measuring and Fostering Non-Cognitive Skills in Adolescence: Evidence from Chicago Public Schools and the OneGoal Program. Working paper, Chapin Hall and University of Chicago.

Kirkeboen, Lars, Edwin Leuven, and Magne Mogstad, 2014. Field of Study, Earnings, and Self-Selection. NBER Working Paper 20816.

Kobrin, J., Patterson, B., Shaw, E., Mattern, K., and Barbuti, S., 2008. Validity of the SAT for predicting first-year college grade point average. College Board Research Report No. 2008-5.

Larranaga O. and D. Contreras D., 2010. Las Nuevas Políticas de Protección Social. United Nations Development Programme, Uqbar, Santiago.

Lee, D., and D. Card, 2008. Regression discontinuity inference with specification error. Journal of Econometrics, 142(2), 655-674.

Lee, D. and T. Lemieux, 2010. Regression Discontinuity Designs In Economics. Journal Of Economic Literature 48, pp. 281-355.

Long, M. C., 2004. Race and College Admissions: An Alternative to Affirmative Action?. Review of Economics and Statistics, 86(4), 1020-1033.

Long, M. C., 2015. Is There a 'Workable' Race-Neutral Alternative to Affirmative Action in College Admissions? Journal of Policy Analysis and Management Volume 34, Issue 1, pages 162-183..

Ludwig, J., and Miller, D. L., 2007. Does head start improve children's life chances? Evidence from a regression-discontinuity design. Quarterly Journal of Economics, 122(1), 159-208.

McCrary, Justin, 2008. Manipulation of the running variable in the regression discontinuity design: A density test, Journal of Econometrics, 142, 698-714.

Mineduc, 2014a. Duracion Real de las Carreras en Chile. Servicio de Informacion de Educacion Superior, Ministerio de Educacion, Gobierno de Chile.

Mineduc, 2014b. Desercion y Reingreso a Educacion Superior en Chile. Servicio de Informacion de Educacion Superior, Ministerio de Educacion, Gobierno de Chile.

Niu, Sunny X. and Marta Tienda, 2010. The Impact of the Texas Top 10% Law on College Enrollment: A Regression Discontinuity Approach. Journal of Policy Analysis and Management, 29(1): 84-110.

OECD and The World Bank, 2009. Reviews of National Policies for Education: Tertiary Education in Chile. Paris: OECD Publishing.

OECD, 2010. PISA 2009 Results: Overcoming Social Background Equity in Learning Opportunities and Outcomes (Volume II).

Pearson Education, 2013. Final Report Evaluation of the Chile PSU.

Porter, J., 2003. Estimation in the regression discontinuity model. Working Paper, Harvard University Department of Economics.

Reyes, Loreto, Jorge Rodriguez, and Sergio S. Urzua. 2013. Heterogeneous Economic Returns to Postsecondary Degrees: Evidence from Chile. NBER Working Paper 18817.

Rothstein, Jesse M., 2004. College performance predictions and the SAT. Journal of Econometrics, 121, pp. 297-317.

Sandel, M., 2009. Justice: What's the Right Thing to do? Farrar, Straus and Giroux, New York. Ch. 7: Arguing Affirmative Action.

Sander, Richard H., 2004. A Systematic Analysis of Affirmative Action in American Law Schools. Stanford Law Review 57(2): 367-483.

Sander, Richard H. and Stuart Taylor Jr., 2012. Mismatch: How Affirmative Action Hurts Students It's Intended to Help, and Why Universities Won't Admit It. Basic Books.

Stratton, Leslie S., Dennis M. O'Toole and James N. Wetzel, 2008. A multinomial logit model of college stopout and dropout behavior. Economics of Education Review 27: 319331.

U.S. Department of Education, National Center for Education Statistics, 2013. Digest of Education Statistics, 2012 (NCES 2014-015), Chapter 3.

Van Der Klaauw, Wilbert, 2002. Estimating The Effect Of Financial Aid Offers On College Enrollment: A Regression-Discontinuity Approach. International Economic Review, Vol. 43, Pp. 1249-1287

White House, 2014. Increasing College Opportunity for Low-Income Students: Promising Models and a Call to Action. The Executive Office of the President.

Zwick, R., and Green, J. G., 2007. New perspectives on the correlation of SAT scores, high school grades, and socioeconomic factors. Journal of Educational Measurement, 44, 23-45.

3.8 Figures and Tables

Table 3.1: Descriptive Statistics

	Validity Sample			Mismatch Sample	
Variable	All Regularly Enrolled [1]	Potentially Displaced [2]	AA Beneficiaries [3]	AA Eligible (Top 5%) [4]	AA Non-eligible (Non-Top 5%) [5]
Pre-university variables					
High School Class Rank (0 to 1)	0.79	0.74	0.98	0.98	0.63
	(0.20)	(0.23)	(0.02)	(0.02)	(0.24)
High School GPA Score	666.6	645.6	712.5	717.9	594.5
	(76.0)	(77.5)	(44.4)	(45.6)	(77.0)
PSU score	648.3	627.8	598.1	598.3	564.7
	(63.9)	(54.2)	(63.5)	(65.0)	(55.2)
Parental Schooling (Years)	14.2	14.0	12.6	12.5	12.8
	(2.9)	(2.9)	(2.7)	(2.8)	(2.8)
Household Income					
Bracket 1 (%) [Lowest]	0.38	0.46	0.63	0.64	0.59
Bracket 2 (%)	0.14	0.12	0.17	0.17	0.19
Bracket 3 (%)	0.28	0.27	0.18	0.18	0.21
Bracket 4 (%) [Highest]	0.20	0.15	0.01	0.01	0.01
Public Health System Use (%)	0.49	0.51	0.74	0.73	0.72
Female (%)	0.48	0.49	0.57	0.59	0.56
Capital Region (%)	0.41	0.39	0.39	0.36	0.24
Scholarship (%)	0.19	0.23	0.31	0.33	0.24
Publicly Funded School (%)	0.73	0.81	1.00	1.00	1.00
High School Quality	298.3	291	258.5	258.6	279.8
	(32.1)	(33.5)	(29.3)	(29.5)	(31.5)
High School Attendance	0.94	0.94	0.95	0.95	0.94
	(0.07)	(0.09)	(0.06)	(0.05)	(0.05)
Number of preferences listed	4.9	5.2	5.6	5.8	5.5
	(1.9)	(1.9)	(1.9)	(1.9)	(2.0)
Preference of enrollment	1.8	1.9	1.9	2.3	2.6
	(1.2)	(1.2)	(1.2)	(1.5)	(1.6)
Year-to-year Persistence					
Year 1	0.83	0.80	0.83	0.82	0.80
Year 2	0.71	0.68	0.70	0.70	0.67
Year 3	0.63	0.62	0.64	0.64	0.59
Year 4	0.62	0.58	0.62	0.62	0.55
Year 5	0.57	0.54	0.60	0.60	0.51
Enrollment					
Continuous (4-5 Years)	0.85	0.83	0.82	0.84	0.74
Stop-out (2-3 Years)	0.12	0.14	0.15	0.12	0.21
Drop-out (0-1 Years)	0.03	0.04	0.04	0.04	0.05
Formal Length (in years)	5.53	5.74	5.75	5.74	5.30
	(1.62)	(1.44)	(1.45)	(1.49)	(1.24)
Observations	18,790	475	466	713	1,765

Notes: Standard deviation for the non-dichotomous variables are in parentheses. Column (1) considers all students regularly enrolled in the same majors as AA beneficiaries. Column (2) is a subset of regularly enrolled students, who would have been displaced by AA beneficiaries as described in the main text. The AA beneficiaries (column 3) correspond to the subset of students that enrolled using the AA policy. Column (4) considers who (i) graduated the previous year from publicly funded schools and took the PSU exam; (ii) applied to flagship universities, (iii) belonged to the first four national quintiles of household income, and (iii) scored below the cutoff of their major of preference (on a 5% band). Those in Column (5) satisfied all same conditions but (iii).

Table 3.2: Validity: Beneficiaries vs. Potentially Displaced Sample

Panel A: 1st Year GPA and Year-to-year Persistence Rates

	GPA	Persistence Rate at:				
	Year 1	Year 1	Year 2	Year 3	Year 4	Year 5
AA Beneficiary ($\widehat{\gamma_m}$)	0.09	0.02	0.02	0.03	0.03	0.06*
	[0.12]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]
	(0.763)	(0.807)	(0.781)	(0.803)	(0.848)	(0.955)
Constant ($\widehat{\beta_m}$)	4.32***	0.80***	0.68***	0.62***	0.58***	0.54***
	[0.08]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]
Observations	324	941	941	941	941	941
1st Year GPA SD	.99					

Panel B: Enrollment Behavior: Continuous, Stopout and Dropout

	Multinomial Logit Model		
	Dropout vs. Continuous	Stopout vs. Continuous	Dropout vs. Stopout
AA Beneficiary	1.070	0.923	1.084
	[0.370]	[0.174]	[0.204]
Observations	941		
chi2(2)	0.236		
Prob > chi2	0.889		

Notes: ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively. Robust standard errors clustered at the high school level are in brackets. The reported coefficients correspond to separate regressions of the dependent variable on the variable 'Best5', defined as 1 for individuals at the top 5% from their high school. All regressions control for gender, high school quality, parental schooling, household income, PSU and GPA scores, and public health use. The persistence rates equals 1 to students that stay in their same pair school-major from year to year. Those that graduated early (or at shorter-duration majors) are counted as persisted. Panel A: all non-parametric specifications use a rectangular kernel. The local-linear bandwidths correspond to 1, 3 and 5 percentiles of class rank. The IK and CV bandwidth represent the Imbens-Kalyanaraman and Cross-Validation method (Ludwig and Miller, 2007) optimal bandwidths, respectively. Panel B specifies linear, quadratic and cubic polynomial functions for $f(rank_i)$, allowing slopes to vary at either sides of the class rank cutoff.

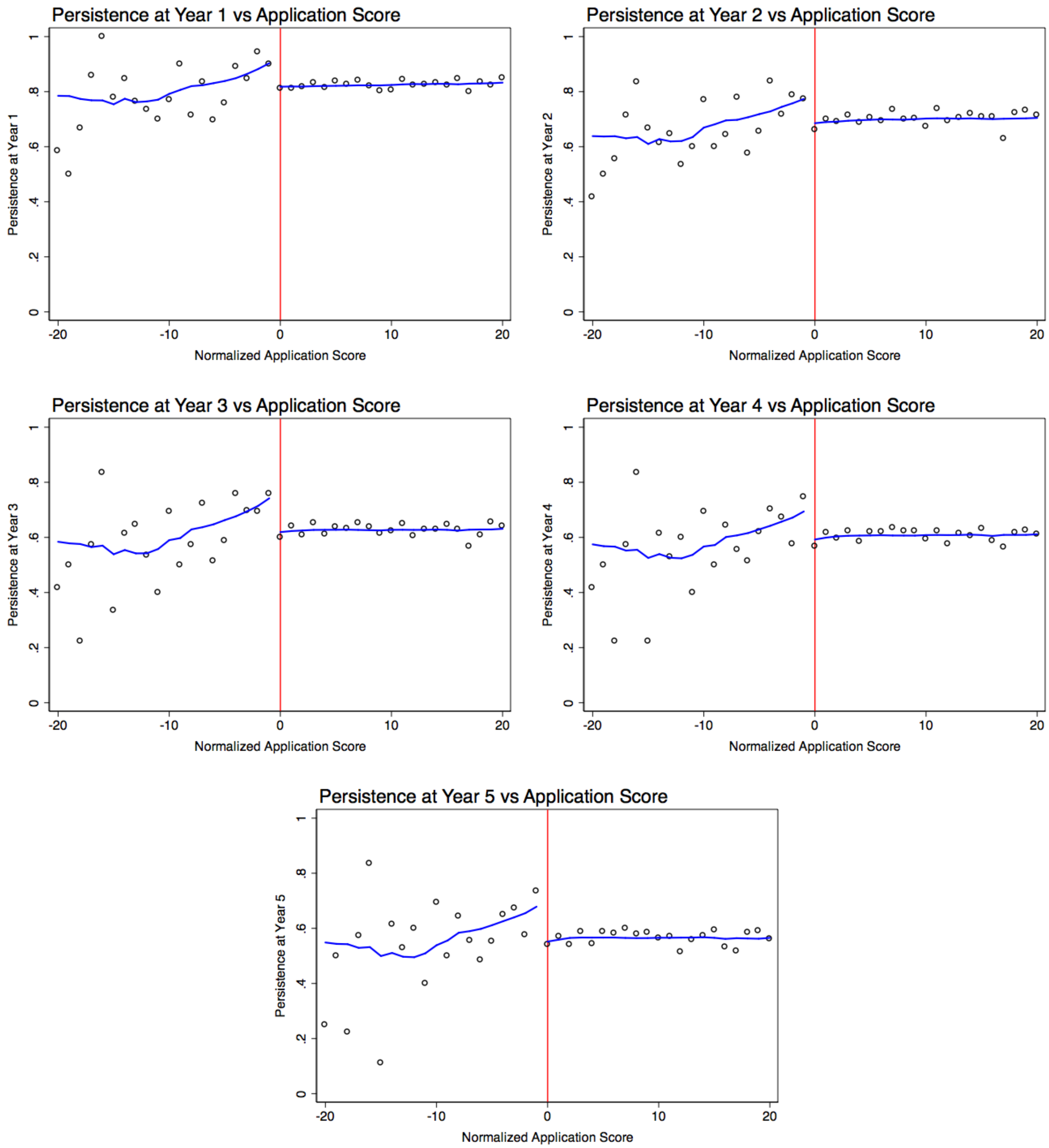


Figure 3.1: Validity Sample: Persistence Rates

Notes: The open circles plot the average values of each variable for application score bins. The solid lines are fitted values from local linear regressions using a rectangular kernel with an optimal cross-validation bandwidth.

Table 3.3: Validity Sample: Persistence Rates (coefficient on AA Beneficiary).

Panel A: AA and Potentially Displaced Students

	Dependent Variable: Persistence Rate at:				
	Year 1	Year 2	Year 3	Year 4	Year 5
AA Beneficiary	0.060*	0.078*	0.088**	0.078*	0.104**
	[0.031]	[0.040]	[0.042]	[0.042]	[0.043]
Observations	941	941	941	941	941

Panel B: Non-Parametric Estimations (AA and all Regularly Enrolled Students)

Local-linear Bandwidth	Dependent Variable: Persistence Rate at:				
(of Application Score)	Year 1	Year 2	Year 3	Year 4	Year 5
Bandwidth=5	0.131***	0.092*	0.135**	0.112*	0.153**
	[0.039]	[0.056]	[0.058]	[0.061]	[0.063]
Observations	3,759	3,759	3,759	3,759	3,759
Bandwidth=10	0.109***	0.107**	0.138***	0.112**	0.140***
	[0.033]	[0.045]	[0.048]	[0.049]	[0.050]
Observations	6,738	6,738	6,738	6,738	6,738
Bandwidth=20	0.074**	0.081**	0.110***	0.089**	0.119***
	[0.027]	[0.035]	[0.037]	[0.037]	[0.038]
Observations	11,322	11,322	11,322	11,322	11,322

Panel C: Parametric Estimations (AA and all Regularly Enrolled Students)

Polynomial Function	Dependent Variable: Persistence Rate at:				
(of Application Score)	Year 1	Year 2	Year 3	Year 4	Year 5
Linear Spline	0.053**	0.058*	0.074**	0.054	0.077**
	[0.024]	[0.032]	[0.034]	[0.034]	[0.035]
Quadratic Spline	0.114***	0.121***	0.163***	0.138***	0.168***
	[0.030]	[0.040]	[0.042]	[0.043]	[0.044]
Cubic Spline	0.116***	0.090*	0.139***	0.109**	0.134**
	[0.037]	[0.050]	[0.052]	[0.055]	[0.056]
Observations	19,256	19,256	19,256	19,256	19,256

Notes: ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively. Robust standard errors clustered at the high school level are in brackets. The reported coefficients correspond to separate regressions of the dependent variable on the variable 'AA Beneficiary', defined as 1 for individuals that enroll thanks to the AA policy. I do not include additional covariates, because the admission system predictability is based solely on the application score. Panel A: all non-parametric specifications use a rectangular kernel. The local-linear bandwidths correspond to 5, 10 and 20 points of the normalized application score. Panel B specifies linear, quadratic and cubic polynomial functions for the application score, allowing slopes to vary at either sides of the class rank cutoff. Panel C reports results from a regression of each dependent variable on the AA beneficiary dummy, for a sample that considers the AA beneficiaries and the group of regularly enrolled students that would have been displaced by them, had the AA policy not consisted in extra slots.

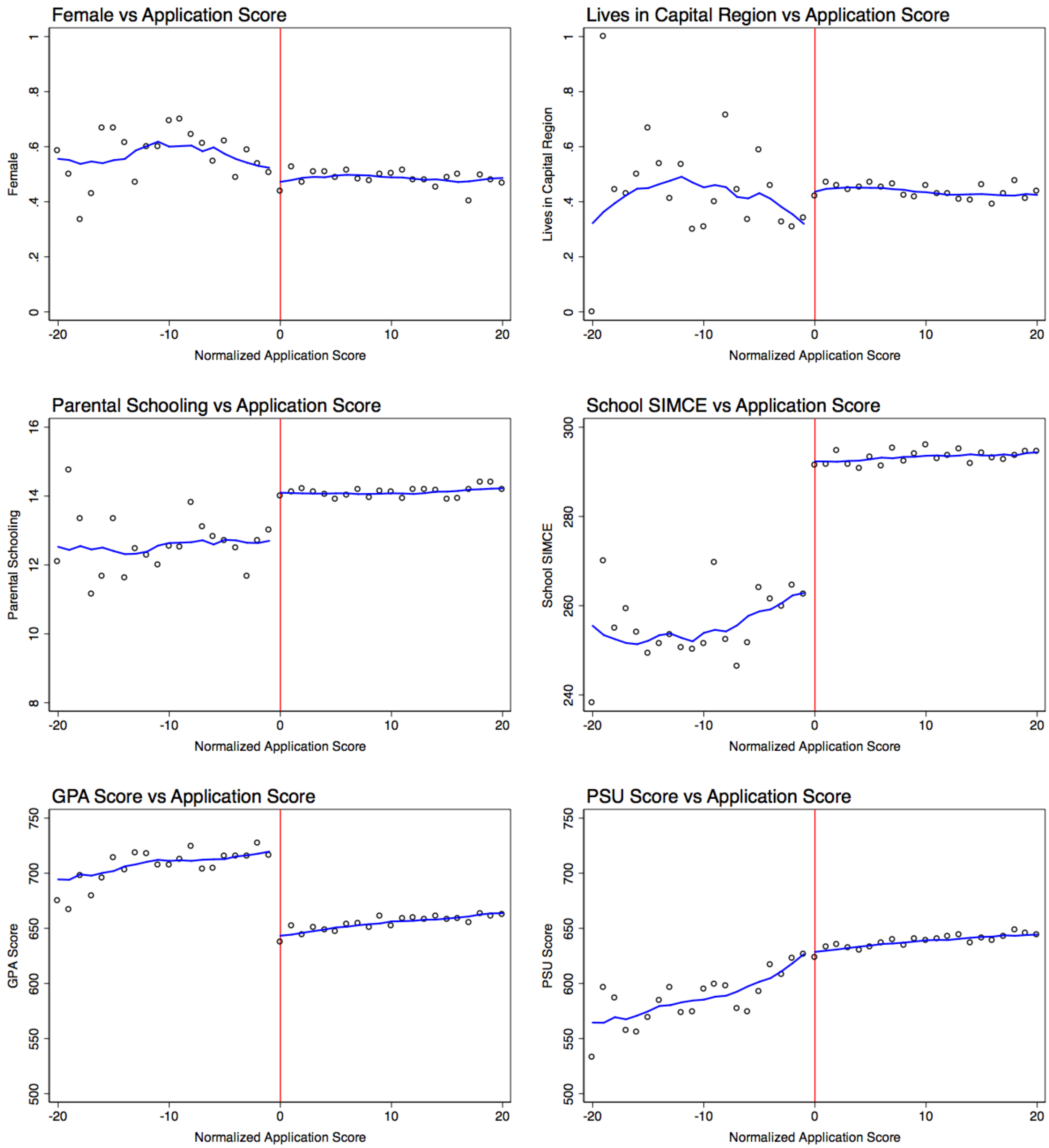


Figure 3.2: Validity Sample: Covariates Differences

Notes: The open circles plot the average values of each variable for application score bins. The solid lines are fitted values from local linear regressions using a rectangular kernel with an optimal cross-validation bandwidth.

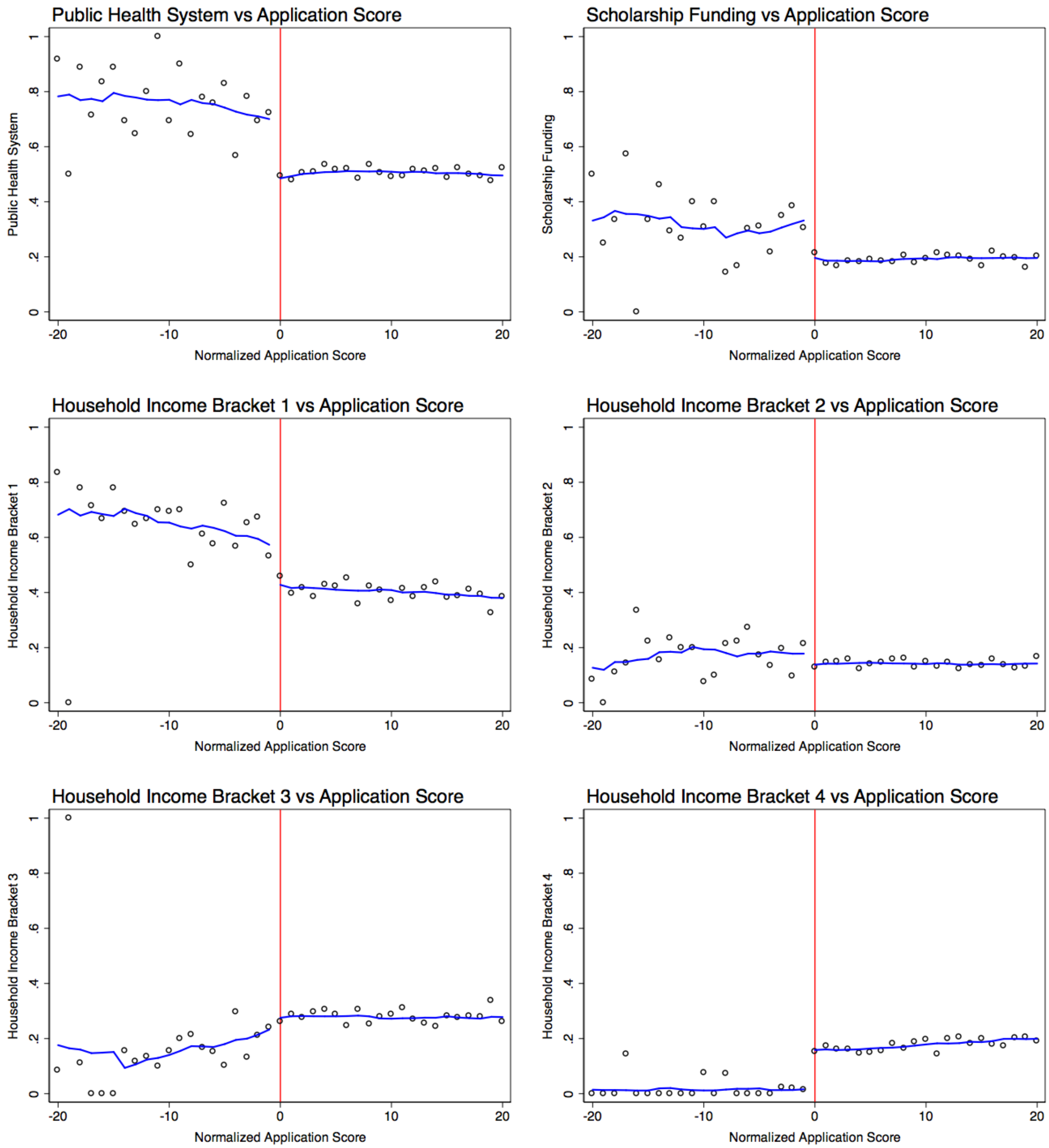


Figure 3.3: Validity Sample: Covariates Differences (continued)

Notes: The open circles plot the average values of each variable for application score bins. The solid lines are fitted values from local linear regressions using a rectangular kernel with an optimal cross-validation bandwidth.

Table 3.4: Validity Sample: Covariates Differences (coefficient on AA Beneficiary).

Panel A: AA and Potentially Displaced Students

	Dependent Variable:										
	Female	HS Quality	Parental Schooling	Income Bracket 1	2	3	4	HS-GPA Score	PSU Score	Public Health	Capital Region
AA Beneficiary	0.048	-31.147***	-1.359***	0.154***	0.056	-0.076**	-0.135***	74.288***	-9.116*	0.221***	0.001
	[0.044]	[2.960]	[0.236]	[0.046]	[0.034]	[0.035]	[0.020]	[5.370]	[5.331]	[0.039]	[0.045]
Observations	900	900	900	900	900	900	900	900	900	900	900

Panel B: Non-Parametric Estimations (AA and all Regularly Enrolled Students)

Local-linear Bandwidth (of Score)	Dependent Variable:										
	Female	HS Quality	Parental Schooling	Income Bracket 1	2	3	4	HS-GPA Score	PSU Score	Public Health	Capital Region
Bw=5 points	0.058	-28.865***	-1.206***	0.123*	0.073	-0.048	-0.148***	74.891***	-1.618	0.230***	-0.125*
	[0.066]	[3.762]	[0.331]	[0.073]	[0.059]	[0.058]	[0.021]	[6.154]	[8.261]	[0.063]	[0.070]
Observations	3,565	3,565	3,443	3,565	3,565	3,565	3,565	3,565	3,565	3,565	3,565
Bw=10 points	0.020	-27.791***	-1.557***	0.160***	0.042	-0.054	-0.149***	77.879***	-1.700	0.210***	-0.136***
	[0.052]	[3.160]	[0.272]	[0.054]	[0.043]	[0.041]	[0.019]	[4.856]	[6.630]	[0.047]	[0.052]
Observations	6,393	6,393	6,142	6,393	6,393	6,393	6,393	6,393	6,393	6,393	6,393
Bw=20 points	0.051	-29.570***	-1.394***	0.168***	0.047	-0.071***	-0.144***	75.943***	-8.526*	0.202***	-0.090**
	[0.040]	[2.535]	[0.210]	[0.041]	[0.032]	[0.032]	[0.014]	[4.104]	[5.115]	[0.036]	[0.042]
Observations	10,719	10,719	10,281	10,719	10,719	10,719	10,719	10,719	10,719	10,719	10,719

Panel C: Parametric Estimations (AA and all Regularly Enrolled Students)

Polynomial Function (of Score)	Dependent Variable:										
	Female	HS Quality	Parental Schooling	Income Bracket 1	2	3	4	HS-GPA Score	PSU Score	Public Health	Capital Region
Linear Spline	0.033	-31.931***	-1.389***	0.181***	0.042	-0.075***	-0.149***	71.964***	-12.586***	0.212***	-0.057
	[0.035]	[2.466]	[0.189]	[0.037]	[0.029]	[0.028]	[0.013]	[3.700]	[4.692]	[0.032]	[0.039]
Quadratic Spline	0.058	-26.452***	-1.315***	0.155***	0.039	-0.049	-0.145***	78.452***	-0.026	0.182***	-0.132***
	[0.047]	[2.921]	[0.246]	[0.048]	[0.038]	[0.038]	[0.015]	[4.487]	[5.919]	[0.043]	[0.048]
Cubic Spline	-0.017	-26.633***	-1.312***	0.131**	0.045	-0.035	-0.141***	72.634***	-0.962	0.182***	-0.151**
	[0.058]	[3.550]	[0.307]	[0.063]	[0.050]	[0.049]	[0.017]	[5.338]	[7.400]	[0.055]	[0.060]
Observations	18,203	18,203	17,331	18,203	18,203	18,203	18,203	18,203	18,203	18,203	18,203

Notes: ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively. Robust standard errors clustered at the high school level are in brackets. The reported coefficients correspond to separate regressions of the dependent variable on the variable 'AA Beneficiary', defined as 1 for individuals that enroll thanks to the AA policy. Panel A: all non-parametric specifications use a rectangular kernel. The local-linear bandwidths correspond to 5, 10 and 20 points of the normalized application score. Panel B specifies linear, quadratic and cubic polynomial functions for the application score, allowing slopes to vary at either sides of the class rank cutoff. Panel C reports results from a regression of each dependent variable on the AA beneficiary dummy, for a sample that considers the AA beneficiaries and the group of regularly enrolled students that would have been displaced by them, had the AA policy not consisted in extra slots.

Table 3.5: Correlation Matrix: Pre-university Skill Measures

	PSU Math	PSU Lang	HS GPA	Class Rank	Attendance
PSU Mathematics	1.00				
PSU Language	0.46	1.00			
HS GPA	0.27	0.25	1.00		
Class Rank	0.11	0.11	0.70	1.00	
Attendance	0.01	-0.05	0.13	0.18	1.00

Notes: N=19,256. The table shows the correlations among five pre-university skill measures. PSU Math and PSU Lang refer to the subscores in Mathematics and Language on the PSU test, respectively. HS GPA corresponds to the cumulative high-school grade average. Class rank indicates the relative position of the students within their high school class. Attendance corresponds to the average annual attendance rate during high-school.

Table 3.6: Correlation Persistence-Measures of Skill

Persistence	PSU Math	PSU Lang	HS GPA	Class Rank	Attendance
Year 1	0.07	0.05	0.07	0.08	0.04
Year 2	0.04	0.05	0.10	0.11	0.05
Year 3	0.00	0.04	0.11	0.12	0.05
Year 4	0.03	0.07	0.13	0.13	0.06
Year 5	0.04	0.05	0.13	0.13	0.06

Notes: N=19,256. The table shows the correlations between persistence rates for years one to five (in the rows), and five pre-university skill measures (in the columns). PSU Math and PSU Lang refer to the subscores in Mathematics and Language on the PSU test, respectively. HS GPA corresponds to the cumulative high-school grade average. Class rank indicates the relative position of the students within their high school class. Attendance corresponds to the annual attendance rate during high-school.

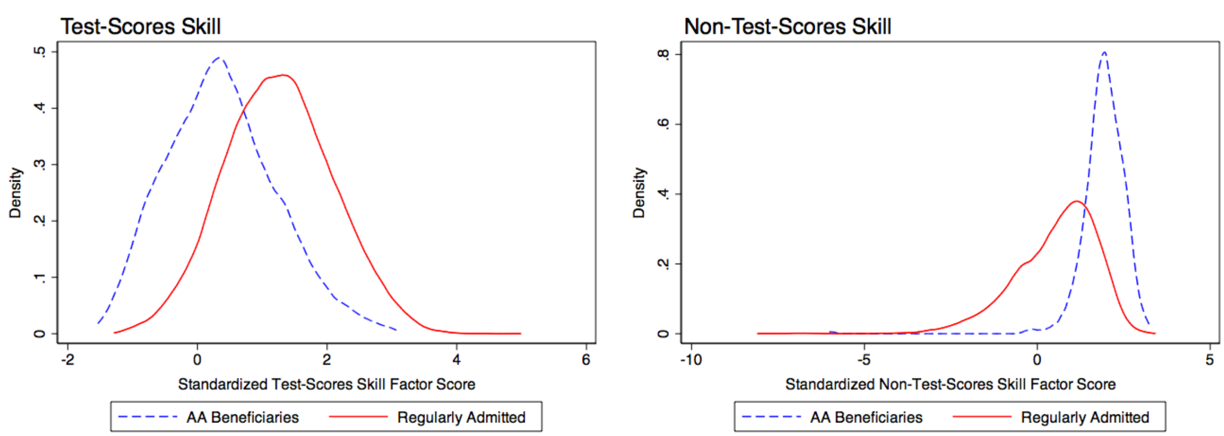


Figure 3.4: Distribution of Skills: $\widehat{\theta}_1$ (left panel) and $\widehat{\theta}_2$ (right panel)

Notes: The left panel shows the distribution of predicted cognitive skill factor scores, based on the subscores from the mathematics and language PSU test. The right panel shows the distribution of predicted non-test-scores skill factor scores, based on the high-school GPA, class rank and attendance rates. The scores have been standardized to have a mean of zero and a standard deviation of one.

Table 3.7: Validity Data: Persistence Rates (coefficient on AA Beneficiary).

	Dependent Variable: Persistence Rate at				
	Year 1	Year 2	Year 3	Year 4	Year 5
Mean Difference ($\widehat{\delta_m}$)	-0.008	-0.007	0.009	0.002	0.025
	[0.018]	[0.022]	[0.023]	[0.023]	[0.023]
Cutoff Difference ($\widehat{\delta_c}$)	0.053**	0.062*	0.077**	0.057*	0.078**
	[0.024]	[0.032]	[0.034]	[0.034]	[0.035]
Observations	19,256	19,256	19,256	19,256	19,256
Regular Students Mean	0.83	0.71	0.63	0.62	0.57
Controlling for Test-Score and Non-test-score skills:					
Mean Difference ($\widehat{\delta_m}$)	-0.017	-0.040*	-0.042*	-0.051**	-0.026
	[0.018]	[0.022]	[0.023]	[0.023]	0.023]
Cutoff Difference ($\widehat{\delta_c}$)	0.025	0.004	0.007	-0.021	-0.001
	[0.024]	[0.032]	[0.034]	[0.034]	[0.035]
Observations	19,256	19,256	19,256	19,256	19,256

Table 3.8: Validity Sample: Persistence Rates (coefficient on AA Beneficiary), controlling for Test Scores and Non-Test Score Skills

Panel A: Non-Parametric Estimations

Local-linear Bandwidth (of Application Score)	Dependent Variable: Persistence Rate at: Year 1	Year 2	Year 3	Year 4	Year 5
Bandwidth=5	0.100***	0.028	0.061	0.027	0.066
	[0.039]	[0.056]	[0.058]	[0.061]	[0.062]
Observations	3,759	3,759	3,759	3,759	3,759
Bandwidth=10	0.073**	0.039	0.059	0.024	0.048
	[0.033]	[0.045]	[0.048]	[0.049]	[0.050]
Observations	6,738	6,738	6,738	6,738	6,738
Bandwidth=20	0.041	0.021	0.040	0.009	0.036
	[0.027]	[0.035]	[0.037]	[0.037]	[0.038]
Observations	11,322	11,322	11,322	11,322	11,322

Panel B: Parametric Estimations

Polynomial Function (of Application Score)	Dependent Variable: Persistence Rate at Year 1	Year 2	Year 3	Year 4	Year 5
Linear Spline	0.025	0.004	0.007	-0.021	-0.001
	[0.024]	[0.032]	[0.034]	[0.034]	[0.035]
Quadratic Spline	0.080***	0.061	0.091**	0.055	0.081*
	[0.030]	[0.040]	[0.042]	[0.043]	[0.044]
Cubic Spline	0.084**	0.031	0.068	0.027	0.049
	[0.037]	[0.050]	[0.052]	[0.055]	[0.055]
Observations	19,256	19,256	19,256	19,256	19,256

Panel C: Potentially Displaced Sample

	Dependent Variable: Persistence Rate at Year 1	Year 2	Year 3	Year 4	Year 5
AA Beneficiary	0.031	0.011	0.015	-0.022	-0.003
	[0.035]	[0.045]	[0.048]	[0.046]	[0.047]
Observations	941	941	941	941	941

Notes: ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively. Robust standard errors clustered at the high school level are in brackets. The reported coefficients correspond to separate regressions of the dependent variable on the variable 'AA Beneficiary', defined as 1 for individuals that enroll thanks to the AA policy. Panel A: all non-parametric specifications use a rectangular kernel. The local-linear bandwidths correspond to 5, 10 and 20 points of the normalized application score. Panel B specifies linear, quadratic and cubic polynomial functions for the application score, allowing slopes to vary at either sides of the class rank cutoff. Panel C reports results from a regression of each dependent variable on the AA beneficiary dummy, for a sample that considers the AA beneficiaries and the group of regularly enrolled students that would have been displaced by them, had the AA policy not consisted in extra slots.

Table 3.9: Expanded Enrollment Opportunities for AA Eligible Applicants

Preference of Admission ($j_{ik}{}^*$)	Preference of Enrollment 1	2	3	4	5	6	7	8	Total
1	4	0	0	0	0	0	0	0	4
2	159	103	0	0	0	0	0	0	262
3	46	45	70	0	0	0	0	0	161
4	18	21	28	28	0	0	0	0	95
5	8	7	9	9	19	0	0	0	52
6	3	4	2	3	6	9	0	0	27
7	3	1	1	4	1	3	11	0	24
8	1	1	3	1	2	3	0	3	14
∅	21	17	16	11	8	1	0	0	74
Total	263	199	129	56	36	16	11	3	713

Eligible, not Beneficiaries (on the diagonal):	247
Eligible, and Beneficiaries (below the diagonal):	466

Notes: The preference of admission ($j_{ik}{}^*$) correspond to the eligible students' best option, given the set of majors A_i where they would have been admitted by the regular system. The preference of enrollment correspond to the preference at which students actually enroll. Among the 713 AA eligible students, 247 do not become beneficiaries, while 466 do.

Table 3.10: Factual and Counterfactual Majors

Student	Enrollment	Major-University Combination		Distance to the Major's Cutoff Score
1	Factual	Medicine	U. de Valparaiso	−12
	Counterfactual	Audiology	U. de Valparaiso	+70
2	Factual	Law	U. de Chile	−10
	Counterfactual	Law	PUC de Valparaiso	+49
3	Factual	Odontology	U. Austral de Chile	−3
	Counterfactual	Nursery	U. de la Frontera	+45

Notes: Each of the three rows indicate the factual and counterfactual major for a particular beneficiary student. The first row presents the case for a student that used the AA policy to enroll in Medicine instead of in Audiology, within the same university (Universidad de Valparaiso). Note that she presented an application score 12 points below the major's cutoff. Had the policy not existed, her best option would have been to enroll in Audiology, for which she presented a score 70 points above the Audiology's cutoff score. The same logic applies for the next cases. The second row illustrates a case where the beneficiary student was able to enroll at a different university (U. de Chile instead of PUC de Valparaiso), within the same major (Law). The third student was able to choose both a different major (Odontology instead of Nursery) and university (U. Austral de Chile instead of U. de la Frontera).

Table 3.11: Factual and Counterfactual Majors Quality

Measure of Quality	Major		Mean
	Factual	Counterfactual	Difference
Prestige Index	5.32	5.11	0.21**
	(0.03)	(0.03)	(0.03)
Expected Wages	2,329.00	2,053.00	276.00**
	(67.95)	(61.16)	(48.39)

Notes: ** difference significant at 5%. Standard errors in parentheses. The 'Que Pasa' Magazine Prestige Index ranges from 1.0 (worse) to 7.0 (best). Expected Wages are monthly quantities, expressed in US Dollars (Year 2008).

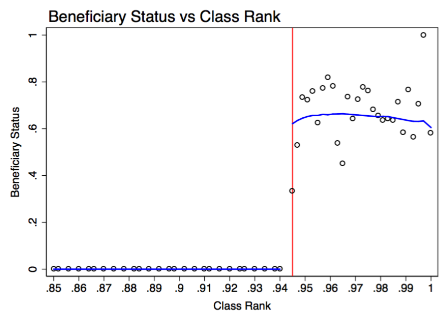


Figure 3.5: Mismatch Sample: Beneficiary Status and Class Rank
Notes: The open circles plot the proportion of AA beneficiaries for class rank bins. The solid lines are local linear regressions of beneficiary status using a rectangular kernel and an optimal cross-validation bandwidth (∼3 percentage points of class rank).

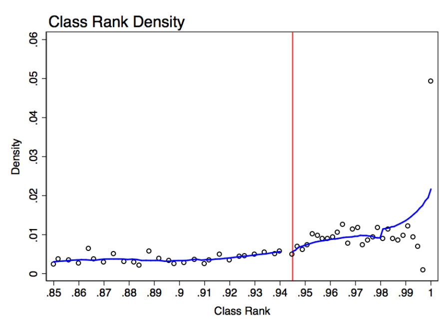


Figure 3.6: Mismatch Sample: Class Rank Density
Notes: The open circles plot the raw density for class rank bins. The solid lines are local linear regressions of density using a rectangular kernel with an optimal cross-validation bandwidth. A McCrary's (2008) test indicates that theres is no sorting in the vicinity of the class rank threshold. The discontinuity estimate is given by .00323, with associated standard errors of .00270, hence non-significant.

Table 3.12: Mismatch Sample: Covariates Smoothness (coefficient on Best 5%).

Panel A: Non-Parametric Estimations

Local-linear Bandwidth (of Class Rank)	Dependent Variable: Female	HS Quality	Parental Schooling	Income Bracket 1	Income Bracket 2	Income Bracket 3	HS-GPA Score	PSU Score	Public Health	Capital Region	Scholarship
Bandwidth= 1 point	0.043	-2.181	0.254	-0.013	-0.047	0.060	-25.553	19.374	0.070	0.214	-0.145
	[0.201]	[13.727]	[1.167]	[0.204]	[0.170]	[0.154]	[15.965]	[24.202]	[0.187]	[0.184]	[0.194]
Observations	137	137	137	137	137	137	137	137	137	137	137
Bandwidth= 3 points	0.160	-2.028	0.169	-0.092	0.028	0.062	-13.919	-2.505	0.079	0.034	-0.070
	[0.099]	[6.338]	[0.567]	[0.096]	[0.078]	[0.079]	[8.577]	[12.736]	[0.088]	[0.095]	[0.091]
Observations	478	478	478	478	478	478	478	478	478	478	478
Bandwidth= 5 point	-0.001	1.048	0.246	-0.045	-0.015	0.056	-7.289	-5.330	0.078	-0.011	-0.017
	[0.076]	[4.843]	[0.457]	[0.077]	[0.062]	[0.058]	[6.452]	[10.018]	[0.068]	[0.072]	[0.069]
Observations	790	790	790	790	790	790	790	790	790	790	790
IK Bandwidth	-0.006	-0.168	0.332	-0.090	-0.037	0.095	-7.945	-2.175	0.061	-0.052	-0.044
	[0.074]	[4.281]	[0.387]	[0.081]	[0.057]	[0.065]	[6.835]	[8.886]	[0.054]	[0.062]	[0.059]
Observations	814	998	1035	693	973	693	723	979	1072	1035	1020
CV Bandwidth	-0.012	1.361	0.460	-0.086	-0.021	0.037	-5.669	-6.804	0.068	-0.017	-0.047
	[0.073]	[4.977]	[0.436]	[0.080]	[0.064]	[0.055]	[6.988]	[9.855]	[0.070]	[0.075]	[0.066]
Observations	827	759	827	723	759	827	693	814	759	723	827

Panel B: Parametric Estimations

Polynomial Function (of Class Rank)	Dependent Variable: Female	HS Quality	Parental Schooling	Income Bracket 1	Income Bracket 2	Income Bracket 3	HS-GPA Score	PSU Score	Public Health	Capital Region	Scholarship
Linear Spline	-0.083*	-2.921	0.026	0.012	0.002	-0.008	5.844	9.216	0.018	0.024	-0.015
	[0.044]	[2.642]	[0.243]	[0.042]	[0.034]	[0.033]	[4.003]	[5.543]	[0.038]	[0.043]	[0.041]
Quadratic Spline	-0.001	-2.421	0.360	-0.109	0.049	0.080	6.730	10.508	-0.010	-0.016	0.014
	[0.070]	[4.093]	[0.367]	[0.069]	[0.055]	[0.057]	[6.328]	[8.078]	[0.060]	[0.068]	[0.064]
Cubic Spline	-0.092	-3.182	0.868	-0.098	0.062	0.053	-8.837	21.035**	0.112	0.029	-0.073
	[0.097]	[5.697]	[0.477]	[0.099]	[0.081]	[0.082]	[9.445]	[10.474]	[0.083]	[0.098]	[0.089]
Observations	2,474	2,474	2,474	2,474	2,474	2,474	2,474	2,474	2,474	2,474	2,474

Notes: ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively. Robust standard errors clustered at the high school level are in brackets. The reported coefficients correspond to separate regressions of the dependent variable on the variable 'Best5', defined as 1 for individuals at the top 5% from their high school. Panel A: the local-linear bandwidths correspond to 1, 3 and 5 percentiles of class rank. The IK and CV bandwidth represent the Imbens-Kalyanaraman and Cross-Validation method (Ludwig and Miller, 2007) optimal bandwidths, respectively. Panel B specifies linear, quadratic and cubic polynomial functions for $f(rank_i)$, allowing slopes to vary at either sides of the class rank cutoff.

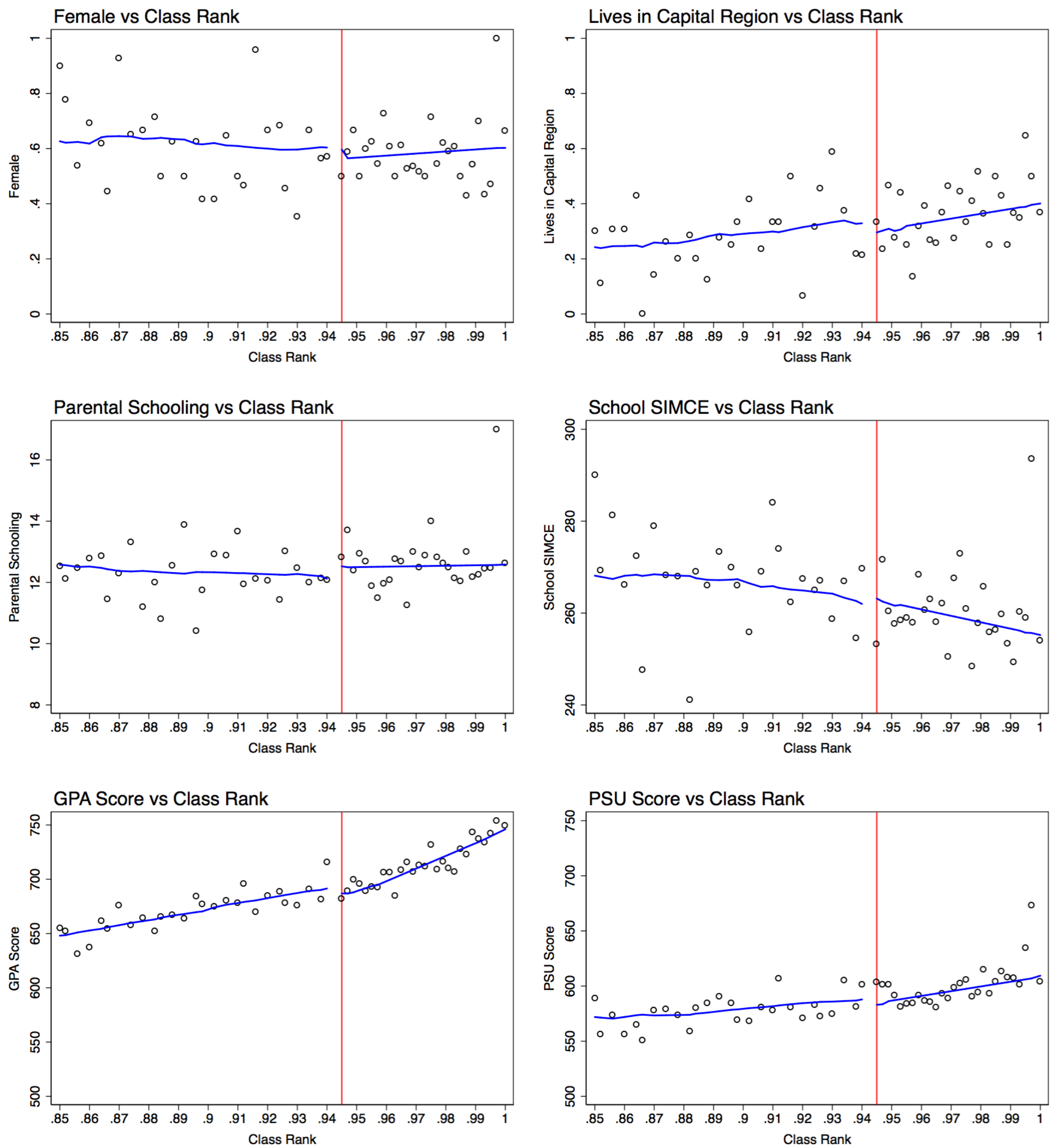


Figure 3.7: Mismatch Sample: Covariates Smoothness

Notes: The open circles plot the average values of each variable for class rank bins. The solid lines are fitted values from local linear regressions using a rectangular kernel with an optimal cross-validation bandwidth.

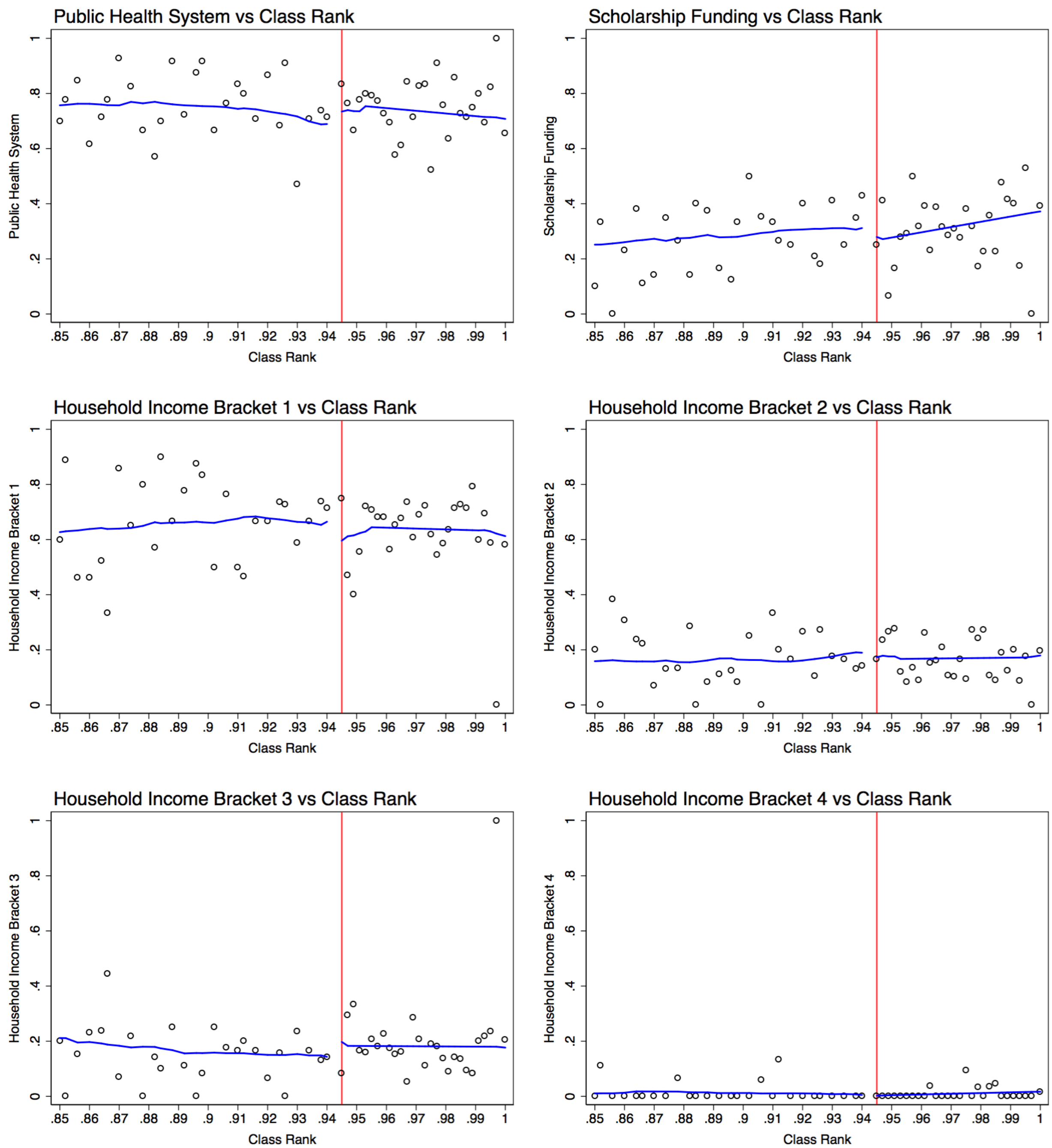


Figure 3.8: Mismatch Sample: Covariates Smoothness (continued)

Notes: The open circles plot the average values of each variable for class rank bins. The solid lines are fitted values from local linear regressions using a rectangular kernel with an optimal cross-validation bandwidth.

Table 3.13: Mismatch Sample: Persistence Rates (coefficient on Best 5%)

Panel A: Non-Parametric Estimations

Local-linear Bandwidth (of Class Rank)	Dependent Variable: Persistence Rate at: Year 1	Year 2	Year 3	Year 4	Year 5
Bandwidth= 1 point	-0.019	0.056	0.313	0.641	0.751
	[0.839]	[0.649]	[0.667]	[0.611]	[0.694]
Observations	124	124	124	124	124
Bandwidth= 3 points	-0.142	-0.170	-0.176	-0.122	-0.191
	[0.153]	[0.160]	[0.170]	[0.189]	[0.191]
Observations	438	438	438	438	438
Bandwidth= 5 point	-0.028	-0.064	-0.080	-0.078	-0.084
	[0.091]	[0.098]	[0.103]	[0.107]	[0.110]
Observations	729	729	729	729	729
IK Bandwidth	0.023	-0.077	-0.044	-0.007	-0.065
	[0.072]	[0.067]	[0.070]	[0.081]	[0.068]
Observations	1015	1144	1083	981	1186
CV Badnwidth	-0.001	-0.061	-0.086	-0.082	-0.051
	[0.089]	[0.096]	[0.100]	[0.104]	[0.114]
Observations	751	751	751	751	699

Panel B: Parametric Estimations

Polynomial Function (of Class Rank)	Dependent Variable: Persistence Rate at: Year 1	Year 2	Year 3	Year 4	Year 5
Linear Spline	-0.002	-0.010	0.035	0.027	0.019
	[0.042]	[0.044]	[0.048]	[0.049]	[0.050]
Quadratic Spline	0.004	0.005	0.041	0.007	0.011
	[0.067]	[0.071]	[0.080]	[0.078]	[0.080]
Cubic Spline	-0.095	-0.107	-0.094	-0.058	-0.016
	[0.089]	[0.099]	[0.108]	[0.108]	[0.110]
Observations	2,404	2,404	2,404	2,404	2,404

Notes: ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively. Robust standard errors clustered at the high school level are in brackets. The reported coefficients correspond to separate regressions of the dependent variable on the variable 'Best5', defined as 1 for individuals at the top 5% from their high school. All regressions control for gender, high school quality, parental schooling, household income, PSU and GPA scores, and public health use. The persistence rates equals 1 to students that stay in their same pair school-major from year to year. Those that graduated early (or at shorter-duration majors) are counted as persisted. Panel A: all non-parametric specifications use a rectangular kernel. The local-linear bandwidths correspond to 1, 3 and 5 percentiles of class rank. The IK and CV bandwidth represent the Imbens-Kalyanaraman and Cross-Validation method (Ludwig and Miller, 2007) optimal bandwidths, respectively. Panel B specifies linear, quadratic and cubic polynomial functions for $f(rank_i)$, allowing slopes to vary at either sides of the class rank cutoff.

Table 3.14: Mismatch Sample: Enrollment Behavior (coefficient on Best 5%)

Panel A: Non-Parametric Estimations

Local-linear Bandwidth (of Class Rank)	Dependent Variable: Dropout Behavior	Continuous Enrollment
Bandwidth= 1 point	-0.406	0.707
	[0.423]	[0.559]
Observations	124	124
Bandwidth= 3 points	-0.045	0.004
	[0.078]	[0.145]
Observations	438	438
Bandwidth= 5 point	-0.052	0.024
	[0.048]	[0.088]
Observations	729	729
IK Bandwidth	-0.021	0.013
	[0.039]	[0.062]
Observations	913	1009
CV Bandwidth	-0.047	0.003
	[0.046]	[0.084]
Observations	751	751

Panel B: Parametric Estimations

Polynomial Function (of Class Rank)	Dependent Variable: Dropout Behavior	Continuous Enrollment
	Dropout	Continuous
Linear Spline	-0.001	0.009
	[0.020]	[0.040]
Quadratic Spline	-0.012	0.006
	[0.030]	[0.066]
Cubic Spline	0.033	-0.099
	[0.041]	[0.097]
Observations	2,404	2,404

Notes: ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively. Robust standard errors clustered at the high school level are in brackets. The reported coefficients correspond to separate regressions of the dependent variable on the variable 'Best5', defined as 1 for individuals at the top 5% from their high school. All regressions control for gender, high school quality, parental schooling, household income, PSU and GPA scores, and public health use. The persistence rates equals 1 to students that stay in their same pair school-major from year to year. Those that graduated early (or at shorter-duration majors) are counted as persisted. Panel A: all non-parametric specifications use a rectangular kernel. The local-linear bandwidths correspond to 1, 3 and 5 percentiles of class rank. The IK and CV bandwidth represent the Imbens-Kalyanaraman and Cross-Validation method (Ludwig and Miller, 2007) optimal bandwidths, respectively. Panel B specifies linear, quadratic and cubic polynomial functions for $f(rank_i)$, allowing slopes to vary at either sides of the class rank cutoff.

Table 3.15: Mismatch Sample: Relative Persistence Rates (coefficient on Best 5%)

Panel A: Non-Parametric Estimations

Local-linear Bandwidth (of Class Rank)	Dep. Variable: Relative Persistence Rate at: Year 1	Year 2	Year 3	Year 4	Year 5
Bandwidth= 1 point	-0.109	-0.155	-0.209	-0.178	-0.206
	[0.142]	[0.149]	[0.152]	[0.151]	[0.184]
Observations	137	137	137	137	137
Bandwidth= 3 points	-0.064	-0.131	-0.042	-0.051	0.038
	[0.075]	[0.079]	[0.081]	[0.082]	[0.094]
Observations	478	478	478	478	478
Bandwidth= 5 point	-0.015	-0.013	0.021	-0.001	0.121
	[0.057]	[0.059]	[0.062]	[0.064]	[0.075]
Observations	790	790	790	790	790
IK Bandwidth	-0.007	-0.029	0.005	-0.006	0.083
	[0.038]	[0.053]	[0.057]	[0.056]	[0.067]
Observations	1405	1035	990	1020	979
CV Bandwidth	-0.017	-0.026	0.000	-0.024	0.139*
	[0.057]	[0.059]	[0.061]	[0.063]	[0.076]
Observations	814	814	814	814	759

Panel B: Parametric Estimations

Polynomial Function (of Class Rank)	Dep. Variable: Relative Persistence Rate at: Year 1	Year 2	Year 3	Year 4	Year 5
Linear Spline	-0.006	-0.015	0.025	-0.007	0.097**
	[0.032]	[0.035]	[0.036]	[0.037]	[0.043]
Quadratic Spline	-0.006	0.005	0.050	0.018	0.142**
	[0.055]	[0.059]	[0.061]	[0.061]	[0.063]
Cubic Spline	-0.117	-0.136*	-0.073	-0.085	0.104
	[0.074]	[0.082]	[0.084]	[0.083]	[0.088]
Observations	2,474	2,474	2,474	2,474	2,474

Notes: ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively. Robust standard errors clustered at the high school level are in brackets. The reported coefficients correspond to separate regressions of the dependent variable on the variable 'Best5', defined as 1 for individuals at the top 5% from their high school. All regressions control for gender, high school quality, parental schooling, household income, PSU and GPA scores, and public health use. I measure relative persistence as the distance in persistence as compared to the mean from their respective major. Those that graduated early (or at shorter-duration majors) are counted as persisted. Panel A: all non-parametric specifications use a rectangular kernel. The local-linear bandwidths correspond to 1, 3 and 5 percentiles of class rank. The IK and CV bandwidth represent the Imbens-Kalyanaraman and Cross-Validation method (Ludwig and Miller, 2007) optimal bandwidths, respectively. Panel B specifies linear, quadratic and cubic polynomial functions for $f(rank_i)$, allowing slopes to vary at either sides of the class rank cutoff.

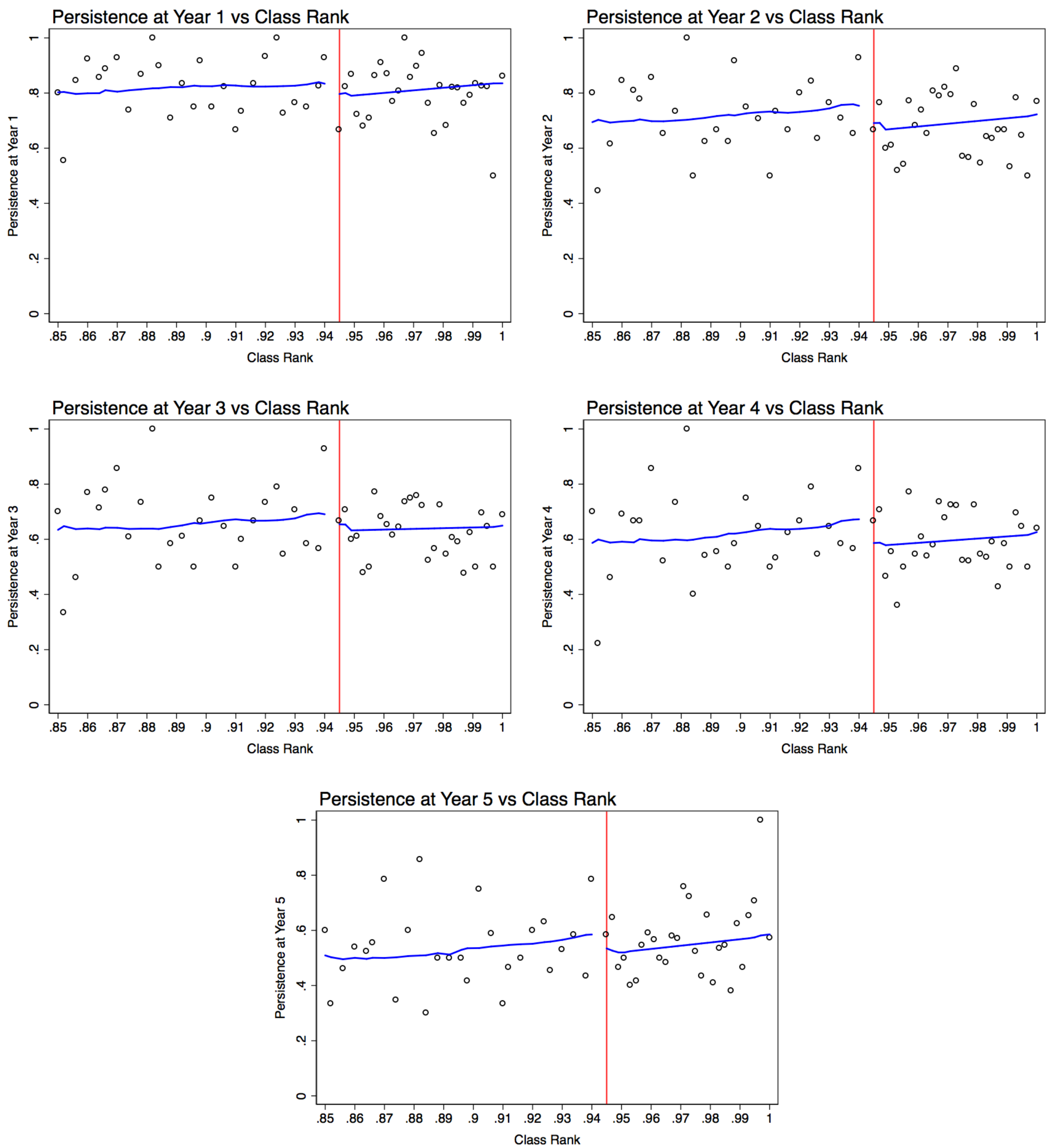


Figure 3.9: Mismatch Sample: Persistence Rates

Notes: The open circles plot the average values of each variable for class rank bins. The solid lines are fitted values from local linear regressions using a rectangular kernel with an optimal cross-validation bandwidth.

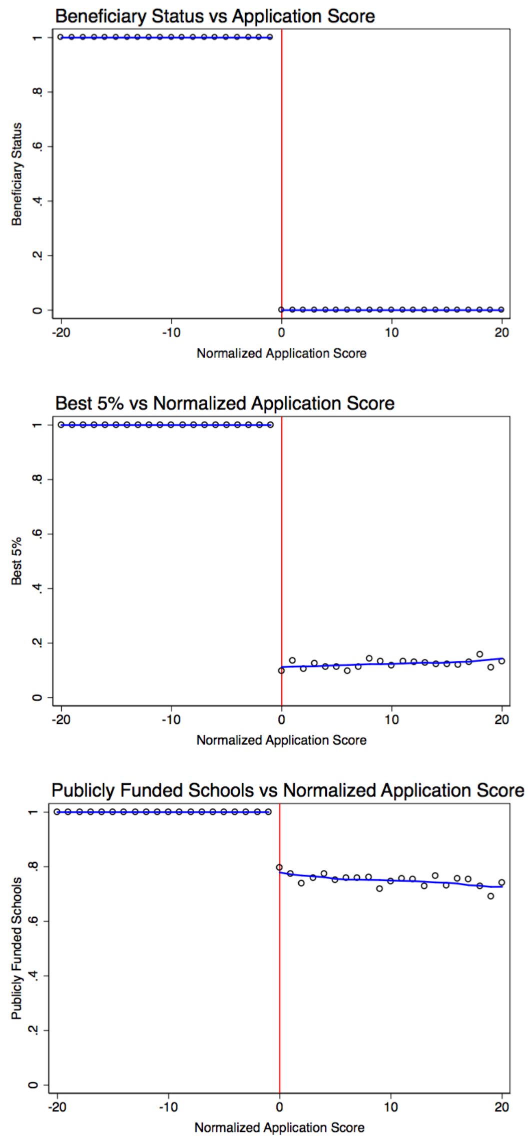


Figure 3.10: Validity Sample: Beneficiary Status Conditions

Notes: The open circles plot the proportion of AA beneficiaries for application score bins. The solid lines are local linear regressions of beneficiary status using a rectangular kernel and an optimal cross-validation bandwidth.

Table 3.16: Descriptive Statistics: Subsample with 1st Year GPA

	AA Beneficiaries	Regularly Enrolled
High School Class Rank (0 to 1)	0.98	0.77
	(0.02)	(0.22)
High School GPA Score	710.6	636.6
	(46.9)	(71.7)
PSU score	598.4	634.2
	(56.1)	(52.5)
Parental Schooling (Years)	12.3	14.4
	(2.8)	(2.9)
Public Health System Use (%)	0.75	0.44
Female (%)	0.56	0.41
Capital Region (%)	0.70	0.72
Publicly Funded School (%)	1.00	0.77
Observations	162	162

Notes: Standard deviation for the non-dichotomous variables are in parentheses. The AA beneficiaries (column 3) correspond to the subset of eligible students that enrolled using the AA policy. The regularly enrolled students comprise all who enrolled according to the regular selection system in the same majors as the AA beneficiaries. High-School GPA corresponds to the cumulative high-school grade average. The persistence rates equals 1 to students that stay in their same pair school-major from year to year. Those that graduated early (or at shorter-duration majors) are counted as persisted.

CHAPTER 4

PERSISTENCE IN THE TRANSMISSION OF EDUCATION: EVIDENCE ACROSS THREE GENERATIONS FOR CHILE

(With P. Celhay)

4.1 Introduction

Most previous studies of intergenerational transmission of human capital are restricted to parents and their children. There is less direct evidence on mobility across multiple generations, which would better answer questions on how opportunities for economic success evolve in the long run. However, very recent theoretical (Solon 2014, 2015; Stuhler 2012) and empirical work is being developed to address this gap.[1] In fact, a rapidly evolving literature is emerging, studying mobility patterns across three or more generations in developed countries.[2]

The results from this new empirical literature consistently show that multigenerational persistence tends to be high compared to the predictions by seminal theoretical models (Becker and Tomes 1979, 1986), which implied that advantages and disadvantages of ancestors tend to disappear rapidly. That older generations –beyond parents– influence children outcomes has important implications for how we interpret intergenerational transmission of human capital. In the absence of direct evidence for more than two generations, the literature previously extrapolated longer term mobility using the available parentchild elasticities. When confronted by the new evidence, this procedure seems to have been mechanically

1. For instance, see the special issue on Inequality Across Multiple Generations (March 2014), based on a recent conference organized by the Panel Study of Income Dynamics in Michigan.

2. Some examples are Chan and Boliver (2013) on class mobility and Long and Ferrie (2013) on occupational mobility, for the UK; Olivetti, Paserman, and Salisbury (2014) on earnings mobility for the USA; Hertel and Groh-Samberg (2014) comparing class mobility between the USA and Germany; Clark (2014), who uses sur- names to study social mobility among a group of eight countries, and Lindahl et al. (2014, 2015) and Havari and Savegnago (2013) on educational mobility for Sweden, and for a group of European countries, respectively.

overestimating longer run mobility patterns.

The evidence found for developed countries leads to question whether similar patterns hold for developing countries. Answering this turns to be of particular importance for the Latin American region, where nations have been systematically characterized by low levels of mobility across two generations as documented in Hertz et al. (2007) and Behrman, Gaviria, and Szekely (2001).[3] Understanding long-run mobility in developing countries would also help to better track the effectiveness of social policies such as compulsory schooling laws and higher education reforms, and to study whether such policies temporarily mask inequalities among families or have lasting effects on individual's relative fortunes and opportunities.

Despite its importance and due to lack of data, this evidence is rare. This paper is one of the first to examine long-run educational mobility for a Latin American country. We use a Chilean data set that enables us to link information on educational attainment for three generations. To study the intergenerational transmission of schooling, we first estimate standard reduced form equations for two and three generations. To this end, we use years of completed schooling for each generation measured in both absolute and relative terms. Our main findings can be summarized as follows. We find the coefficient between grandparents' education and parents' education to be 0.68, which decreases to 0.46 for parents and children. This indicates that absolute mobility increased across generations, that is, that individuals have consistently attained higher levels of education compared to their ancestors. However, relative educational mobility has remained stable. An increase in one standard deviation (SD) in grandparental education is associated with 0.59 of a SD for parents,[4] while the same estimate for parents and children is 0.58. This suggests that the schooling attainment ranking changes little between generations.

3. There is extensive previous research documenting intergenerational mobility on educational attainment across different countries, contexts, and periods. Good examples are Bowles (1972), de Broucker and Underwood (1998), and Jger (2007).

4. As a benchmark consider Hertz et al. (2007), which is a 50-year trends study across 42 countries that showed that 1 SD difference in parental education corresponded to a global average educational difference of 0.4 SDs in the subsequent generation.

Importantly, our estimates show that grandparental education matters even after taking parental education into account. Accordingly, the standard two-generation extrapolation method under-predicts schooling persistence. This is consistent with the evidence found for developed countries, albeit when estimated at higher levels of immobility. As described by Solon (2015), there are many different processes capable of explaining these higher order coefficient estimates. Grandparents may contribute beyond what is mediated through the parent, either directly via interaction with their grandchildren (Zeng and Xie 2014), or indirectly via genetic transmission.[5] There may also be group effects associated with race or ethnicity (Nybom and Stuhler 2014; Oreopoulos and Page 2006) or even geo- graphical differences (Chetty et al. “Is the United States Still a Land of Opportunity?” 2014, “Where is the Land of Opportunity?” 2014) affecting the mobility patterns.

Our data allow us to examine how some of these factors relate to educational mobility. In particular, we address the relationship among the educational mobility patterns and institutional changes, geographical differences, and gender differentials. We first study how the intergenerational transmission of schooling has evolved over time. A birth cohort analysis shows that, in a context of high persistence, educational mobility has moderately increased over time. Auxiliary data from the Chilean Census suggest that such patterns may be attributable to changes in the distribution of education generated by institutional reforms, such as compulsory schooling laws. We then perform a geographical analysis that reveals that the national average educational persistence hides great heterogeneity within the country. Mobility tends to be higher in areas outside the Metropolitan Region (MR), which is the richest and most populated in the country. In addition, more mobile regions have lower levels of income inequality. These facts are in line with the findings for the USA by Chetty et al. (“Where is the Land of Opportunity?” 2014) and for comparisons across countries such as Corak (2013), suggesting that place of birth is important for upward mobility in

5. Quoting Solon (2014), “ ... manifestations of family genetic traits can skip a generation’ (which) is another way of opening up the possibility of a positive coefficient for grandparents’ status.”

Chile. The next set of results show interesting gender differentials in the association between three generations of the same family, consistent with the work for the USA by Olivetti, Paserman, and Salisbury (2014). Children learn gender roles from their parents, and it may be the case that children mimic other same-sex family members. Our findings indicate that both grandparents matter for grandsons while only grandmothers influence granddaughters' education. Paternal grandmothers are more influential than maternal ones suggesting that gender plays a role in explaining schooling persistence.

In a nutshell, our findings show high persistence of schooling across three generations and suggest different mechanisms that may explain such multi-dimensional phenomena. The rest of this document proceeds as follows. Section 4.2 introduces and characterizes our data in the context of the educational system of Chile and presents our methodology. Section 4.3 describes our basic estimates and the potential avenues through which mobility patterns persist. Section 4.4 concludes.

4.2 Data and Methodology

We use the first wave of the Longitudinal Social Protection Survey (LSPS 2002), a representative survey of the Chilean population affiliated with the pension system.[6] The survey was mainly designed to collect information on individuals' employment and social security history. It also collects information for both parents and all children of the interviewee including age and education. This is a particular advantage of using the LSPS compared to other household surveys that only collect data on children currently living in the interviewee's household, which may lead to sample selection bias (Francesconi and Nicoletti 2006). The LSPS 2002 original sample consists of 17 246 respondents drawn from a sampling frame

6. The following waves use the same individuals with data updated in 2004, 2006, and 2009. We use data from 2002 since we would not gain any additional information from other waves and lose a substantial number of observations due to sample attrition. On a related note, the sampling frame may under-represent lower income households, working women, and workers with informal jobs. However, as we detail in Appendix 1, our sample does not differ greatly in terms of average schooling attainment when compared to a nationally representative sample.

of approximately 8.1 million individuals. We restrict our analysis to living respondents who have at least 1 child of 23 years old or older at the moment of the survey.[7] Our working sample size consists of 4529 respondents born around 1944 (G2 generation), with information on their parents (G1 generation),[8] born around 1925, and all of their offspring (G3 generation), born around 1968. Figure 4.1 shows the distribution of birth years by generation. Generation 3 is restricted to be at least 23 years old, so its distribution is truncated at birth year 1980.

The following rows display information for G3, the child generation. There are a total of 12 648 observations in this generation, 51% of them are male, and they are 36 years old on average. Regarding their schooling, they completed 11.3 years of education, and 27% of them continued on to higher education. When comparing education by gender, both sons and daughters achieve the same years of schooling on average. By a simple comparison of means, Table 4.1 illustrates a steady increase in education over generations. This is the case in both years of schooling and, more sharply, in higher education attendance. Next, we will contextualize these schooling increments within Chile's institutional changes.

4.2.1 Descriptive Statistics

Table 4.1 presents descriptive statistics for our sample.[9] The first panel shows the average, SD, and number of observations for a set of variables regarding the generation of grandpar-

7. This is a standard procedure in the literature as carried out by Behrman, Gaviria, and Szekely (2001) and Hertz et al. (2007). Moreover, evidence for Chile shows that schooling accumulation after 23 years old is very low (Nunez and Miranda 2011). This restriction excludes 10 955 individuals whose children could still be completing their schooling. See Appendix 1 for the full detail regarding the imposed sample restrictions.

8. 8. This is an approximate year of birth for the grandparents' generation. The precise information is not available in the survey for all of G1 individuals.

9. Our sample is one in which the survival rate of the second generation is equal to 1 by constructionalive respondentswith a fraction of grandparents dead at the moment of the interview. Hence, the data do not allow us to study the influence of grandparents on grandchildren in the absence of parents. Similarly, the survival rate for those in the third generation is also equal to 1, because we are keeping respondents with at least one son/daughter over 23 years old.

ents (G1). These are variables reported by each interviewed individual about her parents.[10] The average years of schooling for grandparents is 4.4, in a context when Chilean legislation had established 4 years of minimum mandatory education (explained in the next subsection).[11] Grandfathers tend to be more educated than grandmothers with 5.1 and 4.1 years of schooling, respectively. Furthermore, only a small percentage (3%) completed more than 12 years of schooling, which we label as having attended higher education institutions in Table 4.1.

The next panel describes our sample of interviewed individuals or G2 generation. The sample is 52% male, who are on average 60 years old and have finished 8 years of schooling. The percentage of individuals with more than 12 years of schooling jumps to 15% in this generation, with no evidence of differences in schooling by gender. In addition, 29% of the second generation attended primary school in the MR of Chile.[12]

The following rows display information for G3, the child generation. There are a total of 12 648 observations in this generation, 51% of them are male, and they are 36 years old on average. Regarding their schooling, they completed 11.3 years of education, and 27% of them continued on to higher education. When comparing education by gender, both sons and daughters achieve the same years of schooling on average. By a simple comparison of means, Table 4.1 illustrates a steady increase in education over generations. This is the case in both years of schooling and, more sharply, in higher education attendance. Next, we will contextualize these schooling increments within Chile's institutional changes.

10. As a check, we use the 1960 Chilean Census to compare the overall schooling distribution for individuals born in the same birth cohorts as the G1 sample. We find that our sample is very similar to the one in the Census 1960.

11. G1 schooling corresponds to the mean schooling of both grandfather and grandmother. If there is a missing value, we only consider the schooling of the non-missing ancestor. In all regression analyses, we control for a missing value indicator.

12. The MR contains the capital of the country, Santiago. By the time of the survey, 40% of the population lived in the MR.

4.2.2 Educational Changes in Chile

We need to provide some context regarding the Chilean educational system to describe the basic features of our data. This will give the reader a better understanding of the results we present in Section 3.

By 1920, the Chilean legislation had established four years of minimum mandatory education, which was increased to six years by 1929. Through 1965, primary education consisted of six years of schooling, while the secondary level required students to complete six additional years of education. The 1965 reform extended primary education to 8 mandatory years, ideally for ages 614, while secondary school required 4 more years of schooling. These reforms had the expected effect on school coverage; by 1970, there was universal coverage (97%) in primary education, while secondary schooling increased sharply from 18% to 50% of the population between 1964 and 1970. Over the next decades, coverage in secondary schooling continued to increase to 65% in 1982 and 77% in 1990 (see Elacqua 2011).

Figure 4.2 shows the distribution of years of schooling for the three generations. G1s have a median of 4 years of schooling, while G2s have 6 years and G3s 12 years. Consistent with the institutional arrangements described above, the schooling distribution shifts to the right with younger generations. It is worth noting the relatively high proportionapproximately 1 out of 4of G1s that did not complete even 1 year of education. In the next generation, this proportion reduces to about 1 out of 20, while in the subsequent generation (G3) almost all individuals had at least some schooling. Both grandparental and parental generations experience a spike at six years of schooling suggesting that the 1929 mandatory primary education law affected both. Meanwhile, G3 features its highest spike at 12 years of schooling. In this generation, around 58% of individuals completed at least secondary education. Even though these cohorts were not subject to any compulsory law mandating 12 years of education, secondary school coverage increased heavily since 1970.[13] One hypothesis that could explain this phenomenon is that the mandate for eight years of schooling in 1965 also

13. That law was implemented much later around 2003 (Congreso Nacional de Chile 2003).

had an effect on completing secondary school, since it lowered the time needed to get a high school diploma from six to four years.[14]

4.2.3 A Graphical Approach to Mobility Across Generations

Our first approach to mobility across generations is shown in Figures 4.3 and 4.4. In the first figure, we plot the average years of schooling of generation t in the y-axis against the years of schooling of generation t1 in the x-axis. In the second figure, we plot the corresponding SD of schooling. The triangle dots correspond to the analysis of G1 and G2, while the circular dots represent the analysis of G3 and G2. The diagonal is a 45-degree line. These graphs are useful for at least two important reasons. First, they provide a notion on how mobility behaves at the bottom and the top of the schooling distribution. Second, the slopes depicted by the dots represent a graphical approximation of the regression coefficients between generations, which we present in the next section.

Figures 4.3 and 4.4 give us a first view of how absolute and relative mobility evolve across generations, respectively.

For G1G2, Figure 4.3 shows high mobility at the bottom of the schooling distribution of G1, since most of the triangular dots for the first 8 years of education in this generation (x-axis) are far above the diagonal in the next generation (y-axis). Individuals continue to have more years of education than their parents until 12 years of school. After that the triangular dots are not statistically distinguishable from the diagonal,[15] which implies that there is less mobility at the top of the schooling distribution. A similar result is shown in the transition of the next two generations, G2G3. The figure shows that there is even higher mobility at the bottom than for G1G2, because circular dots are further away from the 45-degree line before 8 years of schooling. The pattern described in Figure 4.4 is similar:

14. Probably, the returns from those additional four years of schooling were much higher than dropping out of school and entering the labor market. The assumption here is that labor market demand and supply forces were playing an important role on students' decision to continue and complete 12 years of schooling.

15. This is mainly due to the fact that we have very few grandparents with education above 12 years of schooling so that the estimates of the education mean for those years are much less precise.

higher mobility at the bottom (dots above the 45 degree line) and lower mobility at the top (dots below). However, while absolute mobility across two generations increases as we move forward in the generational spectrum in Figure 4.3, it remains basically unchanged in Figure 4.4 when we use measures of relative mobility. In the latter, both G1G2 and G2G3 depict the same pattern. This suggests that even though younger generations attain higher levels of education than their ancestors, their relative position in the schooling distribution does not change significantly.

4.2.4 Econometric Specifications

Our econometric specifications follow standard descriptive analyses of mobility between adjacent generations, based on the estimation of a reduced form equation derived from the microeconomic model in Becker and Tomes (1979, 1986). We first estimate a linear regression of years of education of generation (t) on years of education of an older generation $(t-s)$in the same family, of the form,

$$S_{it} = \beta_0 + \beta_1 S_{i,t-s} + f(age_{it}, age_{i,t-s}) + \xi_{it} \tag{4.1}$$

where s [1,2 and f (ageit,ageits) summarizes the fact that we include a quadratic function for the age of each generation in the regression. In this setting, the slope β_1 is a measure of immobility. It can be interpreted as a measure of absolute mobility, as it indicates how an additional year of education in generation t s changes education for generation t. If S_{it} defined relative to its distribution, then the coefficient β_1 represents a measure of relative mobility, indicating how the education attainment ranking changes from one generation to another.

As discussed in Section 4.1, a multitude of processes may rationalize multigenerational mobility patterns. We follow Solon (2014) and trivially extend his "income mobility" model to a three generations "educational mobility" setup.[16] Utility-maximizing parents are concerned about the welfare of their children, and now the process describing how endowments are related over time is allowed to be AR(2). This opens the possibility of grandparent contribution beyond mediated contributions through parents. As we derive in Appendix 2, the model collapses to the following reduced form equation:

$$S_{it} = \beta_0 + \beta_1 S_{i,t-1} + \beta_2 S_{i,t-2} + f(age_{it}, age_{i,t-s}) + \xi_{it} \tag{4.2}$$

which previous researchers have estimated for developed countries (for instance, Behrman and Taubman 1985, for the USA, or more recently, Lindahl et al. 2015, for Sweden). In specification (2), β_1 reflects the influence of parental education on children's education, conditioning on grandparental education. Likewise, our parameter of interest β_2 reflects the association between grandparents' and children's education, taking into account the parental factor. We are interested in testing the null hypothesis $H_0 : \beta_2 = 0$. If rejected, it provides evidence of higher order levels of persistence in educational outcomes. In the next section, we present baseline results from the estimation of Equations (4.1) and (4.2). Then, we extend our analysis to study how mobility patterns have evolved over time, how they vary with geographical differences, and whether gender plays a role in the persistence of educational mobility over three generations.

4.3 Results

4.3.1 Basic Results

In this section, we present our basic estimates of intergenerational mobility using both absolute and relative measures. Table 4.2 exhibits the estimated transmission coefficients for years of education across the three generations under study. Estimates in columns (1)(3) show the results from the bivariate regression (4.1), and column (4) shows the result from

16. The microeconomic model extended for three generations is derived in Appendix 2.

estimating equation (4.2).

The association between educational outcomes for parents (G2) and their children (G3), as given in column (2) shows that there is a persistence of about 0.46 in absolute terms. In other words, one additional year of school for parents is related to about half of another year for their children. This is in the same order of magnitude as those found for Chile by Celhay, Sanhueza, and Zubizarreta (2010) and Nunez and Miranda (2011), and similar to the magnitudes reported by the World Bank in Ferreira et al. (2013). In relative terms, an increase in one SD in parental education4.7 years of schooling in our sampleis associated with 2.2 additional years of schooling in the next generation. This represents approximately 0.58 of a SD, as shown by the standardized coefficient in parentheses in column (2).

An interesting result, shown in column (1), is the high persistence in schooling between grandparents and parents. An additional year of education in the grandparent generation (G1) is related to 0.68 years of schooling in the next generation (G2). In relative terms, one additional SD in G1equivalent to 4.1 more years of schoolingis associated with 3.2 additional years of schooling for G2 or 0.59 of a SD. These results indicate that absolute mobility increased Chile (0.68 to 0.46 years), while relative schooling mobility increased only moderately from G1G2 (0.59 SD) to G2G3 (0.58 SD). This is consistent with the slopes depicted in Figures 3 and 4.

These findings also show that if we compute mobility across three generations by the standard two-generation exponentiation procedure, we would be under-predicting intergenerational persistence in educational attainment over three generations for both absolute and relative measures. Suppose, as is the typical case, that we only had the association between educational outcomes for parents (G2) and their children (G3) as shown in column (2). According to the extrapolation by exponentiation method,[17] we would predict a fairly low coefficient for the association between G3 and G1, given by $0.46^2 = 0.21$. However,

17. Consider $S_{it} = \beta_0 + \beta_1 S_{i,t-1} + \xi_{it}$. If we assume the exact same process for the past generation, then $S_{it-1} = \beta_0 + \beta_1 S_{i,t-2} + \xi_{it-1}$. Replacing the second in the latter, we get $S_{it} = \alpha_0 + \alpha_1 S_{i,t-2} + \varepsilon_{it}$, where $\alpha_0 = \beta_0 + \beta_1$, $\alpha_1 = {\beta_1}^2$, and $\varepsilon_{it} = \beta_1 \xi_{it-1} + \xi_{it}$.

the estimated coefficient 0.39 presented in column (3) is about twice as large. Note that although less severely, this also happens for the relative mobility measure.

Is this grandparental influence acting only through the parents' education? If so, we would expect the coefficient on grandparents' education to be zero when we estimate Equation 4.2. However, column (4) shows that, even after conditioning on parental education, the grandparent factor does not fade out, rejecting the null hypothesis $H_0 : \beta_2 = 0$. In other words, there is a persistent association between G1 and G2, despite the fact that there is one generation or an average of 50 years between the births of G1 and G3.

Lindahl et al. (2015) estimate similar specifications for Sweden, which provides a benchmark to interpret the coefficients for Chile described above. For cohorts that were born in similar years such as ours, their standardized estimated coefficients of Equation (2) are 0.267 for parents (0.499 in our specification) and 0.110 for grandparents (0.120 in our specification), respectively. Interestingly, even in a highly mobile country such as Sweden, the grandparental factor has an independent (from the parent) direct association when predicting grandchildren's years of education. Why do we get this prevalence of schooling over more than one generation? The next subsections of results explore possible avenues through which the transmission of human capital persists over more than one generation.

4.3.2 Changes Over Time: Institutional Channel

We extend the baseline equations to study how mobility among generations has changed over time, and then relate it to the trends in mobility and institutional changes Chile has experienced. First, we construct a variable for the birth cohort of G3. Table 4.3 shows the distribution of birth cohorts grouped into five-year periods.[18] To test for changes in educational mobility across time, we interact G2's education in Equation 4.1 and G1's education in Equation 4.2, with a binary indicator for each cohort group in Table 4.3. In addition, in

18. We pool observations born between 1920 and 1944 due to sample size restrictions and use this cohort as our base group.

the regressions of children's education on grandparents' education, we always condition on parents' years of education, unless stated otherwise.

The results are shown in Table 4.4. The first column shows the estimated parameters from Equation 4.1, relating parents' years of education (G2) to children's years of education (G3), controlling for cohort effects. The negative sign on the interaction coefficients and the fact that they are increasing in absolute value for younger cohorts are indicative of a consistent pattern of increasing mobility between generations G2 and G3 over time[19] The second column shows the results for the same regression, but now for grandparents (G1) and children (G3), controlling for parental education and cohort effects. Similarly, the pattern over time displays an increasing trend in mobility between grandparents and their grandchildren.

A visual test of differential trends is shown in Figure 4.5, where we plot the coefficients from Table 4.4. Interestingly, we see a relatively steeper trend for cohorts born between 1950 and 1964 in both cases, after which the differences from one cohort to another are less obvious. This is more evident when we look at the relation between grandparents' and children's education.[20]

One explanation for such change may be attributed to institutional rearrangements. In 1965, Chile implemented a compulsory schooling law that established mandatory attendance to primary school that, at the time, corresponded to eight years of schooling.[21] Oreopoulos and Page (2006), for the UK, and Nybom and Stuhler (2014), for Sweden, show that compulsory laws have a significant effect on the distribution of schooling for the affected generation. In particular, one may expect compulsory laws to increase the mean and shrink the variance of schooling for the affected cohort, as compared to the precedent cohort not

19. The F-test for the joint significance of the interaction terms is at the bottom of the table.

20. 20. An analog analysis using relative measures of mobility finds the same results (see Figure 4.9).

21. In this same period, to attend the increase in demand as a result of the compulsory schooling law, there was an important expansion of government spending in school infrastructure, teacher training, and food provision in schools, among others. See Gobierno de Chile (1962) and Schiefelbein (1976) for references and anecdotal evidence.

affected by the law.

We show that this is the case for Chile using data from the 2002 Census. Figure 4.6 plots the mean and SD of years of education for each of the G2 cohorts in our sample. The figure illustrates that for birth cohorts that were potentially affected by the compulsory law, that is, those born after 1950, the mean shiftsslightly, and the SD decreases dramatically until 1961. This period coincides with the time of the acceleration in the downward trend shown in Figure 4.5, suggesting that the correlation of G3 education with G2 or G1 education decreases as compulsory laws reduce the dispersion of children's educational attainment. Hence, the data support that compulsory educational laws represent a driving mechanism for upward mobility.

4.3.3 Is it Location? Geographical Differences

Recent studies for other countries have shown that intergenerational mobility varies substantially across geographical areas.[22] For instance, Chetty et al. ("Is the United States Still a Land of Opportunity?" 2014, "Where is the Land of Opportunity?" 2014) find that, in the USA, more racially and income segregated areas as well as higher income inequality areas have lower levels of mobility, and that the quality of local school supply in the K-12 system is positively correlated with mobility. Regarding Chile, there is important geographical variation in terms of school quality, local labor markets, culture, economic activity, and income inequality (Agostini and Brown 2007). In the northern regions, the main economic activities are commerce and mining, while the economy in southern regions is concentrated in agriculture. In addition, households in northern regions have higher than national average household income, while the south has more impoverished households.[23] These facts motivate us to study how intergenerational educational mobility differs across the country.

22. See Corak (2013) for a cross-country variation.

23. See Contreras (2001) for an analysis of the evolution of poverty by Regions between 1990 and 1996. See Gobierno de Chile (2013) for Regional differences in poverty rates and average household income in year 2012.

Historically, Chile has been geographically divided into 13 different regions,[24] with the MR being the most populated followed by the 8th and 5th Regions. These three regions represent about 62% of the country's population, and their main cities are considered as the three great urban centers of Chile. Our data allow us to study changes in intergenerational mobility by grandparents' place of origin. Ideally, one would have the birthplace to test for differences by origin. Given that the survey we use lacks this information, we approximate grandparents' place of origin by the region in which G2 attended primary school. One important question remains: how much did regions differ when G2 was growing up? To address this issue, we do the following. The median birth cohort of parents is 1940. Hence, using the Chilean Census of 1960, we can study regional differences in different out- comes by the time the median G2 was 20 years old. In particular, for each region, we compute the percentage of people who attended less than primary school, the percentage of households without sewage and piped water, and the concentration of industries as reported by employed respondents in the Census. These are shown in Table 4.10. All in all, the patterns of regional differences explained above are similar to the ones in 1960.

In the following, we test for whether the coefficients found in Table 4.2 between G1 and G2, and the ones between G1 and G3, differ by the place where G2 attended primary school. We concentrate on differences from all regions with respect to the MR, that is, the capital. We hypothesize that the MR may have presented better economic opportunities at the time, and hence this comparison would help to test whether intergenerational patterns vary with local characteristics.

In particular, we construct a binary indicator for each region according to G2's primary school and interact each of these indicators with G1's education level. We regress the years of schooling of G2 against that of G1, its interaction with the regional indicators, main effects, and all of the controls included in our baseline results. The coefficient of each interaction shows the differences in the education correlations for those in G2 who attended

24. 24. Since 2007, the number increased to 15.

primary school in these regions as opposed to the MR. We perform the same analysis using G3 schooling as the dependent variable.

The results are shown in Table 4.5. Each coefficient is in deviation from the MR, the omitted category.[25] The first column shows the differences across regions in schooling mobility between grandparents and parents (G1G2). In the MR, an additional year of grandparents' education is associated with 0.63 additional years of schooling for the parents. The interactions show that the relation between grandparents' and parent's education is lower for 7 out of the remaining 12 regions. In particular, it shows significantly higher mobility in Regions 5 and 3 and lower mobility in Region 4 with respect to the MR. The cross-regional differences are jointly significant.

Grandparents may invest in their children thinking about the welfare of the grandchildren generation, among many other reasons. However, their direct influence in the success of grandchildren can be offset or complemented by parent's actions. We run Equation 4.2 to test for whether there are significant cross-regional differences in mobility between grandparents (G1) and their grandchildren (G3), using the region in where the parent (G2) went to primary school. Column (2) in Table 4.5 shows the results of estimating Equation 4.2 with the regional interactions. There still is significant geographical heterogeneity; however, there is a less consistent cross-regional pattern as the one shown previously.

The results in Table 4.5 mirror those in Figure 4.3 where we document higher educational mobility at the bottom of the schooling distribution. In fact, previous studies for Chile find that there is high mobility at the bottom and low mobility at the top of the distribution for different economic outcomes (see Torche 2005). Thus, we would expect to see higher mobility in regions outside the MR since those regions would have been the most disadvantaged for G1. The fading effect in upward mobility outside the MR, when we study mobility from G1 to G3, may be due to the fact that G2 sorted into different regions of the country, so locational differences for their offspring do not follow the same pattern.

25. We group all observations using the former divisions, which consisted of 13 regions in total.

The relation between intergenerational mobility and income inequality has started to be addressed in the literature, mainly due to the work by Corak (2013). He reports a high correlation between intergenerational elasticities of income and income inequality measured by the Gini coefficient using data from 13 developed countries. In addition, Chetty et al. ("Where is the Land of Opportunity?" 2014) construct measures of inter-generational elasticities in income and income inequality for all commuting zones in the USA to study geographical patterns of intergenerational mobility, finding a strong negative correlation between income inequality and upward mobility. We carry out a similar analysis here, applied to intergenerational educational mobility. Figure 4.7 exhibits, in the y-axis, the correlation between the region-specific mobility coefficients from Table 4.5 plotted against regional Gini coefficients, in the x-axis.[26] The first panel shows the relation between the Gini coefficient and the schooling coefficients for schooling of grandparents and parents (G1G2), and between schooling of grandparents and years of schooling of their grandchildren (G1G3). The second panel shows a similar analysis using the probability of attending higher education as the outcome for children.[27] Both panels show a positive relation between downward mobility (a higher coefficient) and inequality across regions in Chile.[28] This is more evident when we use the probability of attending higher education as the dependent variable.[29]

We believe that these results are important in shedding light on how geographical differences matter for upward educational mobility. Differences in education may be driving cross-regional differences in inequality that could be explained by local institutions, characteristics of residents, local labor markets, or other factors. Tracking outcomes at local

26. The different size of the regions is considered by the diameter of each circle or triangle. The biggest is the MR.

27. We run our regressions using the probability of having more than 12 years of schooling as the dependent variable. Higher education is an interesting variable to study since access to higher education is more related to family background given its higher costs, while earlier education may be more inclusive and may grow mechanically with local school supply expansion and compulsory schooling laws.

28. As a reminder, higher coefficients in the regressions are interpreted as lower mobility. Hence, the graphs are showing a positive relation between downward mobility and inequality.

29. We performed the same analysis using relative measures of mobility and the same results were found (see Figure 4.10).

geographical levels is a first step to study, for instance, the importance of the design of decentralized policies in highly centralized countries, as is the case in most of Latin America.

4.3.4 Does Sex Matter? Gender Differentials

Gender roles are related to social norms and cultural attitudes (Farree and Vella 2013; Moen, Erickson, and Dempster-McClain 1997), which may play a role in explaining schooling persistence. Children learn gender roles from their parents, and it may be the case that children mimic other family members such as grandparents. Two-generational mobility studies have found that parents and children of the same gender have stronger ties when measured by the intergenerational educational transmission compared to cross-gender relationships (Fernandez, Folgi, and Olivetti 2004; Olivetti and Paserman 2013). In this section, we present novel evidence of the role of gender in explaining educational mobility across three generations.

The nature of our data allows us to study patrilineal and matrilineal lineages (Chan and Boliver 2013; Hertel and Groh-Samberg 2014), that is, lineages composed by paternal grandfathers, fathers, and sons, and maternal grandmothers, mothers, and daughters, respectively. Figure 4.8 describes such lineages. For instance, if the survey respondent was the father (mother) of the household, we have information on paternal (maternal) grandparents and children.[30] Therefore, we can take a finer step than in previous analyses and estimate gender-specific intergenerational coefficients.[31]

Tables 4.6 and 4.7 present the results for the patrilineal and matrilineal estimations, respectively. Columns (1) and (2) from Table 4.6 exhibit coefficients using the son's years of education as the dependent variable and father's and grandfather's education as explanatory

30. From the 4529 respondents, 2395 are men and 2134 are women. Following the patrilineal lineage, the male respondents have 7029 children, 3647 sons and 3382 daughters. The number of sons with complete information on both their grandmother's and grandfather's schooling, parental education, and age is 2577. There are 2414 daughters with the same data. Regarding matrilineal lineage, the 2134 women report having 5698 children older than 23 years of age. Of them, 2827 are sons and 2871 are daughters. The number of them with complete information is 2333 for both sons and daughtersby chance the same number.

31. By restricting the sample to specific genders in G1, G2, and G3, our sample size diminishes markedly. Regrettably, this prevents us from performing lineages analysis by cohorts given the low number of observations available for the combination gender-generations-cohorts.

variables. The first result is that an additional year of grandfather's education is significantly associated with 0.090 school years for the grandson after we take paternal education into account. When we include grandmother's education in column (2), the paternal grandfather's coefficient remains significant, but the grandparental effect appears to be divided in similar magnitudes between the two. The relation of grandmother's and grandfather's education to the grandson's education is statistically indistinguishable from each other.

Column (3) shows a similar result for granddaughters. An additional year of schooling for grandfathers is associated with 0.085 years of education for granddaughters. However, once we control for grandmother's education, the results show that the grandparental influence is due to only grandmothers. An additional year of grandmother's schooling is associated with 0.159 additional years of schooling for granddaughters, while the effect of grandfathers is not distinguishable from zero. This suggests that while both patrilineal grandparents matter for the grandsons' education, granddaughters' years of schooling are mostly related to their grandmother's education.

An analogous analysis is carried out for matrilineal lineages in Table 4.7. Column (1) shows that an additional year of grandmother's education is associated with 0.079 years of grand- daughter's education after we condition for the mother's education. This effect holds after we include the grandfather's education, which appears to be non-significant. Columns (3) and (4) replicate these analyses for sons. While there is a significant coefficient for grandmother's schooling on grandsons' education, it vanishes when we control for grandfathers.

These results suggest that grandsons benefit from both grandparents within the patrilineal lineage, while the maternal grandfather has the strongest influence for the matrilineal lineage. Meanwhile, the association between grandmother and granddaughters is significant for both lineages, but its magnitude is markedly higher within the patrilineal lineage.[32] That is, women born to low-educated grandmothers and high-educated grandfathers have lower education than granddaughters born to higher educated grandmothers and lower educated

32. Results are similar when considering the relative measure of mobility as shown in Tables 4.11 and 4.12.

grandfathers. This evidence supports the notion that having more educated same-sex ancestors matters more for women and suggests that gender-related social roles may be passed along generations within families.

4.4 Concluding Remarks and Discussion

This paper contributes to a recent and increasing literature studying intergenerational mobility across multiple generations. It is one of the first to address long-run educational mobility for a developing country, complementing the new wave of American and European studies that go beyond two generations to understand the evolution of inequality. In particular, we present novel evidence on educational mobility patterns across three generations in Chile. We study intergenerational transmission of schooling by estimating standard reduced form equations for two and three generations. We first find that there is higher mobility at the bottom and lower mobility at the top of the schooling distribution of the precedent generation for both absolute and relative mobility measures. This holds for transitions from grandparents to parents and parents to children. We also find that absolute mobility increased across generations, that is, that individuals have consistently achieved higher levels of education compared to their ancestors. This can be a reflection of the rise in economic growth and mandatory schooling reforms experienced by Chile. Meanwhile, relative educational mobility has remained stable, which suggests that the schooling attainment ranking has not changed significantly from one generation to another.

An important result from this paper is the finding of persistence in education between grandparents and grandchildren even after conditioning for parental education. This result illustrates that, even in a context of high immobility, the transmission between grand- parents and grandchildren is under-predicted by the standard simple extrapolation of a two-generational regression coefficient. This finding is consistent with the evidence found for developed countries (Lindahl et al. 2015).

We then extend this analysis aiming to get a better understanding on the different mechanisms through which human capital (im)mobility persists; in particular, we examine the relationship among educational mobility patterns and institutional changes, geographical differences, and gender differentials. Evidence shows that upward mobility in education has increased with younger cohorts, and that such changes may be attributable to compulsory schooling laws (as found by Oreopoulos and Page 2006 and Nybom and Stuhler 2014 for developed countries). Our results also suggest that there is important heterogeneity in mobility across regions in Chile, and that regional inequality and mobility are negatively correlated, consistent with what has been recently found for the USA, by Chetty et al. ("Is the United States Still a Land of Opportunity?" 2014, "Where is the Land of Opportunity?" 2014) and by Corak (2013) in a cross-country comparison. This result sheds light on the importance of parents' place of origin in upward mobility. Finally, a gender lineage analysis indicates that while grandsons benefit from the education of both grandparents, grandmothers influence strongly granddaughters' education levels. This suggests that gender-related models may be transmitted within the family even across generations, which is similar to recent findings by Olivetti, Paserman, and Salisbury (2014) for the USA.

A vast new literature is being developed to study correlations between multiple generations in different outcomes such as education, income, social class, and occupation. Our results contribute to this literature by presenting the first evidence for a Latin American country. We understand that surveys like the one we use in this paper are being implemented in other nations from the region. This will surely boost Latin American research on the topic in the near future. Our findings indicate that beyond nuclear family, family members like grandparents matter for educational mobility. Therefore, studying how grandparents influence their grandchildren can benefit the understanding of long-run mobility. Challenges for future work are clearly directed in understanding the role of grandparents and to pin down the causal channels of persistence. Still, the evidence so far indicates that family back- ground effects can be longer lasting than previously believed, affecting the endowments and idiosyncratic capabilities of children. This needs to be taken into account by policies aimed at improving human development and reducing inequality.

4.5 References

Agostini, Claudio, and Phillip Brown. 2007. "Desigualdad geografica en Chile." Revista de Analisis Economico- Economic Analysis Review 22 (1): 333.

Becker, Gary S., and Nigel Tomes. 1979. "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility." Journal of Political Economy 87 (6): 11531189.

Becker, Gary S., and Nigel Tomes. 1986. "Human Capital and the Rise and Fall of Families." Journal of Labor Economics 4 (3, Part 2): 139.

Behrman, Jere R., Alejandro Gaviria, and Miguel Szekely. 2001. "Intergenerational Mobility in Latin America." Economia 2 (1): 131.

Behrman, Jere, and Paul Taubman. 1985. "Intergenerational Earnings Mobility in the United States: Some Estimates and a Test of Becker's Intergenerational Endowments Model." The Review of Economics and Statistics 67 (1): 144151.

Bowles, Samuel. 1972. "Schooling and Inequality from Generation to Generation." The Journal of Political Economy 80 (3): 219251.

Bravo, David, Sankar Mukhopadhyay, and Petra E. Todd. 2010. "Effects of School Reform on Education and Labor Market Performance: Evidence from Chile's Universal Voucher System." Quantitative Economics 1 (1): 4795.

Celhay, Pablo, Claudia Sanhueza, and Jose R. Zubizarreta. 2010. "Intergenerational Mobility of Income and Schooling: Chile 19962006." Revista de Analisis Economico-Economic Analysis Review 25 (2): 4363.

Chan, Tak Wing, and Vikki Boliver. 2013. "The Grandparents Effect in Social Mobility: Evidence from British Birth Cohort Studies." American Sociological Review 78 (4): 117.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility." American Economic Review 104 (5): 141147.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." Quarterly Journal of Economics 129 (4): 26332679.

Clark, Gregory. 2014. The Son also Rises: Surnames and the History of Social Mobility. Princeton, NJ: Princeton University Press.

Congreso Nacional de Chile. 2003. “Reforma Constitucional que establece la Obligatoriedad y Gratuidad de la Educacion Media.” Technical report, Biblioteca Nacional del Congreso de Chile.

Contreras, Dante. 2001. “Economic Growth and Poverty Reduction by Region: Chile 199096.” Development Policy Review 19 (3): 291302.

Corak, Miles. 2013. “Income Inequality, Equality of Opportunity, and Intergenerational Mobility.” The Journal of Economic Perspectives 27 (3): 79102.

De Broucker, Patrice, and Kristen Underwood. 1998. “Intergenerational Education Mobility: An International Comparison with a Focus on Postsecondary Education.” Education Quarterly Review 5 (2): 3051.

Elacqua, Gregory. 2011. “Breve historia de las reformas educacionales en Chile (1813-presente).” Unpublished manuscript. IPP, Universidad Diego Portales.

Farre, Lidia, and Francis Vella. 2013. “The Intergenerational Transmission of Gender Role Attitudes and its Implications for Female Labour Force Participation.” Economica 80 (318): 219247.

Fernandez, Raquel, Alessandra Fogli, and Claudia Olivetti. 2004. “Mothers and Sons: Preference Formation and Female Labor Force Dynamics.” The Quarterly Journal of Economics 119 (4): 12491299.

Ferreira, Francisco HG, Julian Messina, Jamele Rigolini, Luis-Felipe Lopez-Calva, Maria Ana Lugo, Renos Vakis, and Luis Felipe Lo. 2013. Economic Mobility and the Rise of the Latin American Middle Class. Washington, DC: World Bank Publications.

Francesconi, Marco, and Cheti Nicoletti. 2006. “Intergenerational Mobility and Sample Selection in Short Panels.” Journal of Applied Econometrics 21 (8): 12651293.

Gobierno de Chile. 1962. “Conferencia Interamericana sobre Educacion y Desarrollo Economico y Social para la America Latina,” Santiago de Chile, technical report.

Gobierno de Chile. 2013. “Resultados de la Encuesta CASEN 2013,” Ministerio de Desarrollo Social, Gobierno de Chile, technical report.

Havari, Enkelejda, and Marco Savegnago. 2013. “Intergenerational Transmission of Human Capital. Evidence from a Three Generation Study.” Working paper presented at the royal economic society annual conference, Bank of Italy.

Hertel, Florian R., and Olaf Groh-Samberg. 2014. “Class Mobility Across Three Generations in the US and Germany.” Research in Social Stratification and Mobility 35: 3552.

Hertz, Tom, Tamara Jayasundera, Patrizio Piraino, Sibel Selcuk, Nicole Smith, and Alina Verashchagina. 2007. “The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends.” The BE Journal of Economic Analysis & Policy 7 (2): 19351682.

Jger, Mads Meier. 2007. “Educational Mobility Across Three Generations: The Changing Impact of Parental Social Class, Economic, Cultural and Social Capital.” European Societies 9 (4): 527550.

Lindahl, Mikael, Marten Palme, Sofia Sandgren-Massih, and Anna Sjogren. 2014. “A Test of the Becker-Tomes Model of Human Capital Transmission Using Microdata on Four Generations.” Journal of Human Capital 8 (1): 8096.

Lindahl, Mikael, Marten Palme, Sofia Sandgren Massih, and Anna Sjogren. 2015. “Long-Term Intergenerational Persistence of Human Capital: An Empirical Analysis of Four Generations.” Journal of Human Resources 50 (1): 133.

Long, Jason, and Joseph Ferrie. 2013. “Intergenerational Occupational Mobility in Great Britain and the United States Since 1850.” The American Economic Review 103 (4): 11091137.

Moen, Phyllis, Mary Ann Erickson, and Donna Dempster-McClain. 1997. “Their Mother’s Daughters? The Intergenerational Transmission of Gender Attitudes in a World of Changing Roles.” Journal of Marriage and the Family 59 (2): 281293.

Nunez, Javier, and Leslie Miranda. 2011. “Intergenerational Income and Educational Mobility in Urban Chile.” Estudios de economia 38 (1): 195221.

Nybom, Martin, and Jan Stuhler. 2014. “Interpreting Trends in Intergenerational Mobility.” Working Paper Series, Swedish Institute for Social Research.

Olivetti, Claudia, and M. Daniele Paserman. 2013. “In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 18501930.” NBER Working Papers 18822, National Bureau of Economic Research, Inc.

Olivetti, Claudia, M. Daniele Paserman, and Laura Salisbury. 2014. “Intergenerational Mobility Across Three Generations in the 19th Century: Evidence from the US Census.” Mimeo, Boston University.

Oreopoulos, Philip, and Marianne E. Page. 2006. “The Intergenerational Effects of Compulsory Schooling.” Journal of Labor Economics 24 (4): 729760.

Schiefelbein, Ernesto. 1976. “Diagnostico del sistema educacional chileno en 1970,” No. 31. Universidad de Chile, Facultad de Ciencias Economicas y Administrativas, Departamento de Economia.

Solon, Gary. 2014. “Theoretical Models of Inequality Transmission Across Multiple Generations.” Research in Social Stratification and Mobility 35: 1318.

Solon, Gary. 2015. “What Do We Know So Far about Multigenerational Mobility?” NBER Working Papers 21053, National Bureau of Economic Research.

Stuhler, Jan. 2012. “Mobility Across Multiple Generations: The Iterated Regression Fallacy.” Working Paper 7072, IZA.

Torche, Florencia. 2005. "Unequal but Fluid: Social Mobility in Chile in Comparative Perspective." American Sociological Review 70 (3): 422450.

Wooldridge, Jeffrey M. 2007. "Inverse Probability Weighted Estimation for General Missing Data Problems." Journal of Econometrics 141 (2): 12811301.

Zeng, Zhen, and Yu Xie. 2014. "The Effects of Grandparents on Children's Schooling: Evidence from Rural China." Demography 51: 599617.

4.6 Appendix

4.6.1 Appendix 1. Data sample restrictions

Data set sample restrictions: for the purpose of our analysis, we restrict the sample in different ways. The original sample consists of 17 246 individuals. We exclude those that are dead at the moment of the interview (937), and have no children that are 23 years old at the moment of the interview (10 955). This leaves 5354 "usable" primary sample members. In addition we have problems of missing data on the first generation (765), missing data on education (11), and age inconsistencies between generations 2 and 3. We drop 71 observations for which the difference in age between generations 2 and 3 is less than 14 years old. We attempt to control for the missing data problem or show robustness of our results by adjusting sampling weights by Inverse Probability Weighting (see Wooldridge 2007). Briefly, we run a probit model of a binary indicator that equals 1 if the information is missing and 0 otherwise on region of residence age and gender. We adjust the initial sampling weights by the predicted probability of this model. We report results with and without this adjustment in Table 4.8, which show little differences and provide us more confidence on the robustness of our findings.

An additional note on the data set is given by Bravo, Mukhopadhyay, and Todd (2010). As they describe, the sampling frame of the 2002 survey consists of individuals enrolled in the social security system for at least 1 month during the 19812001 time period, which included individuals who in 2002 were working, unemployed, out of the labor force, receiving pensions, or deceased (in which case the information was collected from surviving relatives). The sample was drawn from a sampling frame of approximately 8.1 million current and former affiliates compiled from official databases (which covers approximately 75% of the population). The sampling frame for the next survey, in 2004, was augmented to include individuals not affiliated with the social security system, so that the sample is representative of the entire Chilean population over the age of 15. In their paper, Bravo, Mukhopadhyay, and Todd (2010) supplemented the 2002 survey with 2004 information. Regrettably, we cannot do the same for our purposes, given that the 2004 wave did not collect information on all of the respondent's offspring education. Therefore, we stick with the 2002 information only.

On a concluding mark, the representativeness of our sample is of those affiliated to the pension system in year 2002 and with at least 1 child of 23 years old or more. To check how the lack of national representativeness was affecting the robustness of our results, we computed the schooling attainments for the 2004 LSPS sample. We found the fraction of individuals with higher education to be lower in the nationally representative sample, but differences were not huge (12% compared to 15% in our sample). In addition, we found the average years of schooling to be eight years of education, as in our 2002 sample.

4.6.2 Appendix 2. The extended economic model of intergenerational mobility

Existing theoretical models of intergenerational transmission of socioeconomic status have strong implications for the association of outcomes across multiple generations of a family. These models, however, do not encompass many plausible avenues for transmission across multiple generations. Solon (2014) extends existing models to encompass some of these avenues and draws out empirical implications for the multigenerational persistence of socioeconomic status.

We present theoretical intergenerational mobility model of schooling, which adds the possibility of grandparental contribution to the child education (building on Solon (2014), and Behrman and Taubman 1985, both in the spirit of Becker and Tomes 1979). The following four equations describe the basics of the model. Parents are budget constrained, so they divide their income into consumption and investment in their child human capital (equation (4.3)); there is a technology that transforms that investment into human capital (schooling) for the child (equation (4.4)); an AR(2) process (hence, allowing grandparental contribution[33] beyond the contribution mediated through the parent) describes how endowments are related over time (equation (4.5)); and human capital (schooling) translates into earnings in the labor market according to equation (4.6).

$$y_{i,t-1} = C_{i,t-1} + I_{i,t-1} \tag{4.3}$$

$$S_{it} = \theta log(I_{i,t-1}) + e_{it} \tag{4.4}$$

$$e_{it} = \delta + \lambda_1 e_{i,t-1} + \lambda_2 e_{i,t-2} + v_{it} \tag{4.5}$$

$$\log(y_{it}) = \mu + pS_{it} \tag{4.6}$$

Parents have an altruistic utility function depending on their own consumption $C_{i,t-1}$ and their children's (expected life time) earnings. This variable, in turn, depends on the investment made by parents $I_{i,t-1}$. Parents maximize utility over consumption and investment, subject to the budget constraint (equation (4.3)).

$$\max_{C_{i,t-1}, I_{i,t-1}} (1-\alpha)\log(C_{i,t-1}) + \alpha \log(y_{it}) \;\; s.t. \;\; (4.3)$$

Also, note that consider lagging (4.6)

$$\log(y_{i,t-1}) = \mu + pS_{i,t-1}$$

and therefore,

$$C_{t-1} = \exp(\mu + pS_{i,t-1}) - I_{t-1}$$

33. As pointed out by Stuhler (2012), Lindahl (2013B) and others, there are other ways to extend the underlying Becker-Tomes model.

Additionally, consider replacing (4.4) in (4.6),

$$\log(y_{it}) = \mu + pS_{it}$$
$$\log(y_{it}) = \mu + p(\theta \log(I_{i,t-1}) + e_{it})$$

and denote $\gamma_1 = p\theta$ the elasticity of the child's income with respect to investment in the child's human capital,

$$\log(y_{it}) = \mu + \gamma_1 \log(I_{i,t-1}) + pe_{it}$$

So if parents are aware of equations (4.4) - (4.6), the objective function can be re-written in terms of $I_{i,t-1}$ as,

$$\max_{C_{i,t-1}, I_{i,t-1}} (1-\alpha)\log C_{i,t-1} + \alpha \log(y_{it}) \;\; s.t. \;\; (4.3)$$
$$\max_{I_{i,t-1}} \;\; (1-\alpha)\log(\exp(\mu + pS_{i,t-1}) - I_{t-1}) + \alpha(\mu + \gamma_1 log(I_{i,t-1}) + pe_{it}) \;\; s.t. \;\; (4.3)$$

with first order conditions given by

$$\frac{1-\alpha}{\exp(\mu + pS_{i,t-1}) - I_{i,t-1}} = \frac{\alpha\gamma_1}{I_{i,t-1}}$$

Hence, the optimal investment in human capital is given by

$$I^*_{i,t-1} = \frac{\alpha\gamma_1}{\alpha(\gamma_1 - 1) + 1}\exp(\mu + pS_{i,t-1})$$

Therefore, replacing in (4.4), we get an equation relating the child's schooling with the parental schooling,

$$
\begin{aligned}
S_{it} &= \theta log(I^{*}_{i,t-1}) + e_{it} \\
S_{it} &= \underbrace{\theta \left[log \left(\frac{\alpha\gamma_1}{\alpha(\gamma_1+1)-1} \right) + \mu \right]}_{=\gamma_0} + \gamma_1 S_{i,t-1} + e_{it} \\
S_{it} &= \gamma_0 + \gamma_1 S_{i,t-1} + e_{it}
\end{aligned}
$$

Still, we need to incorporate the human capital endowment process information (e_{it}). Since we are allowing for grandparental contribution -beyond the parental one-, we are going to get a reduced form that links the child's education with the grandparent schooling.

Let's lag the equation from last line in one and two periods, and premultiply them by λ_1 and λ_2, respectively

$$
\begin{aligned}
S_{it} &= \gamma_0 + \gamma_1 S_{i,t-1} + e_{it} \\
\lambda_1 S_{i,t-1} &= \lambda_1\gamma_0 + \lambda_1\gamma_1 S_{i,t-2} + \lambda_1 e_{i,t-1} \\
\lambda_2 S_{i,t-2} &= \lambda_2\gamma_0 + \lambda_2\gamma_1 S_{i,t-3} + \lambda_2 e_{i,t-2}
\end{aligned}
$$

Replacing these last two equations in (4.5),

$$
\begin{aligned}
e_{it} &= \delta + \lambda_1 e_{i,t-1} + \lambda_2 e_{i,t-2} + v_{it} \\
e_{it} &= \delta + \lambda_1 S_{i,t-1} - \lambda_1\gamma_0 - \lambda_1\gamma_1 S_{i,t-2} + \lambda_2 S_{i,t-2} - \lambda_2\gamma_0 - \lambda_2\gamma_1 S_{i,t-3} + v_{it}
\end{aligned}
$$

Replacing the expression for e_{it} in the equation we just lagged, we get,

$$S_{it} = \gamma_0 + \gamma_1 S_{i,t-1} + e_{it}$$
$$S_{it} = \gamma_0 + \gamma_1 S_{i,t-1} + \delta + \lambda_1 S_{i,t-1} - \lambda_1\gamma_0 - \lambda_1\gamma_1 S_{i,t-2} + \lambda_2 S_{i,t-2} - \lambda_2\gamma_0 - \lambda_2\gamma_1 S_{i,t-3} + v_{it}$$

Rearranging,

$$S_{it} = \underbrace{\delta + \gamma_0(1-\lambda_1-\lambda_2)}_{\beta_0} + \underbrace{(\gamma_1+\lambda_1)}_{\beta_1} S_{i,t-1} + \underbrace{(\lambda_2-\gamma_1\lambda_1)}_{\beta_2} S_{i,t-2} + \underbrace{v_{it} - \gamma_1\lambda_2 S_{i,t-3}}_{\xi_{it}}$$

So, our reduced form equation looks like

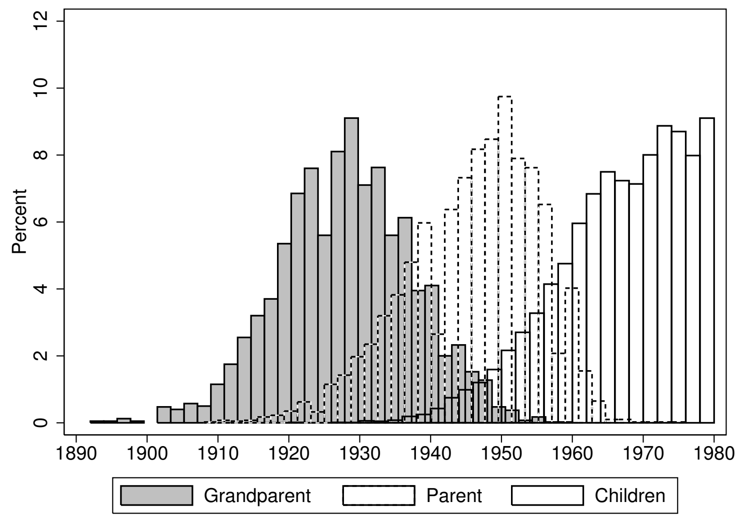
$$S_{it} = \beta_0 + \beta_1 S_{i,t-1} + \beta_2 S_{i,t-2} + \xi_{it} \tag{4.7}$$

Note that the coefficient for grandparental schooling may be positive because of the incorporation of a grandparental contribution to the child's genetic/cultural inheritance. In particular, the grandparental coefficient is positive if our structural parameters satisfy $\lambda_2 - \gamma_1\lambda_1 > 0$. This would imply that grandparental contributions to genetic or cultural inheritance are important enough to more than compensate a parental "bad draw" endowment effect (Solon, 2013)[34].

34. Another point to raise is that the structural parameters from the model (as in the original Becker-Tomes model) are identified since we have three equations and five unknowns $(\delta, \gamma_0, \gamma_1, \lambda_1, \lambda_2)$. Fixing, say, δ and γ_0, we could "identify" the rest of the *more* relevant parameters γ_1, λ_1 and λ_2.

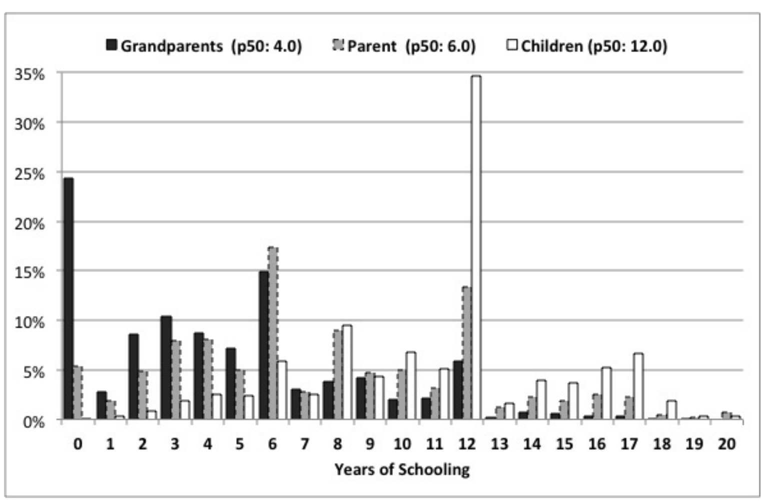
4.7 Figures and Tables

Figure 4.1: Birth cohorts by generation



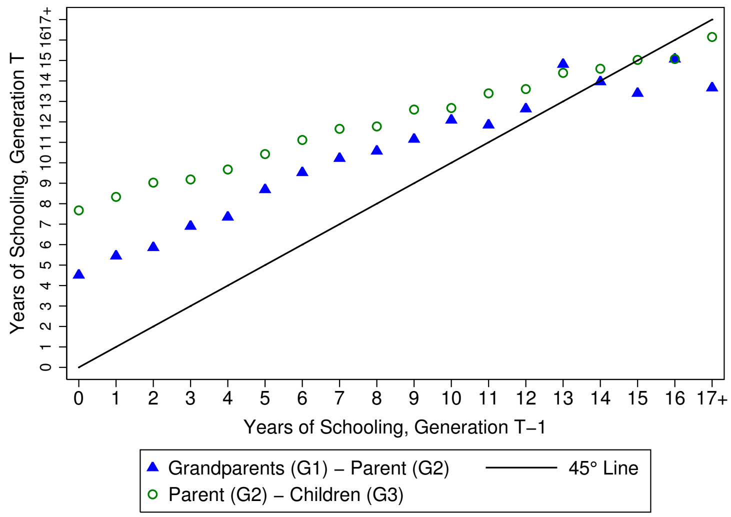
Notes: This figure shows the distribution of the years of birth by generation. Generation 3 is restricted to be 23 years old or older, so its distribution is truncated at birth year 1980. For the G1 generation, the precise information is not available in the survey for all of G1 individuals.

Figure 4.2: Distribution of schooling across generations



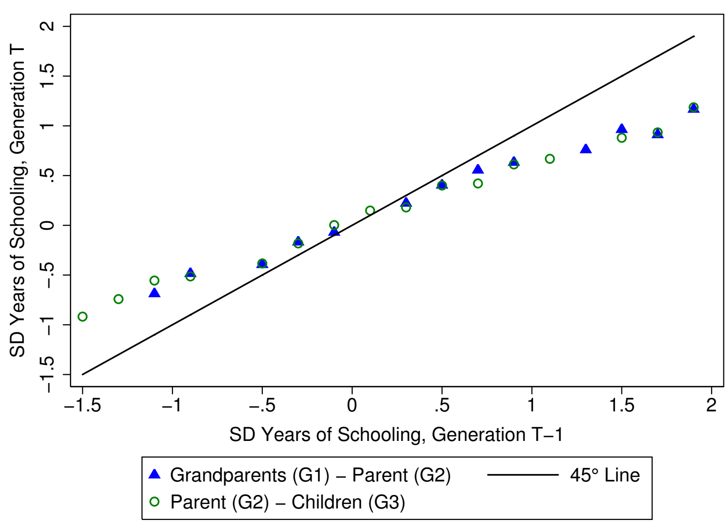
Notes: This figure shows the distribution of years of schooling for the three generations we study. For G1 years of schooling, we use the average using education of both grandparents, or the one that is available when either the grandfather or grandmother information is missing.

Figure 4.3: Absolute mobility across generations



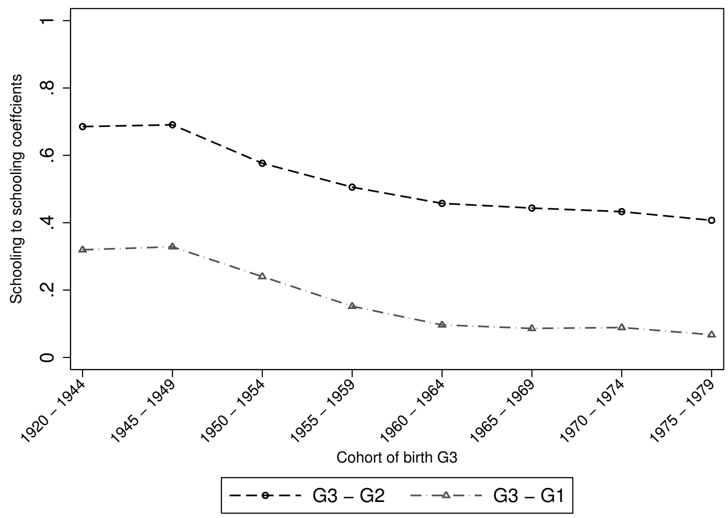
Notes: This figure shows the average years of schooling of generation t in the y-axis, against the years of schooling of generation t1 in the x-axis. The triangle dots correspond to the analysis of G1 and G2, while the circular dots represent the analysis of G3 and G2. The diagonal is a 45-degree line.

Figure 4.4: Relative mobility across generations



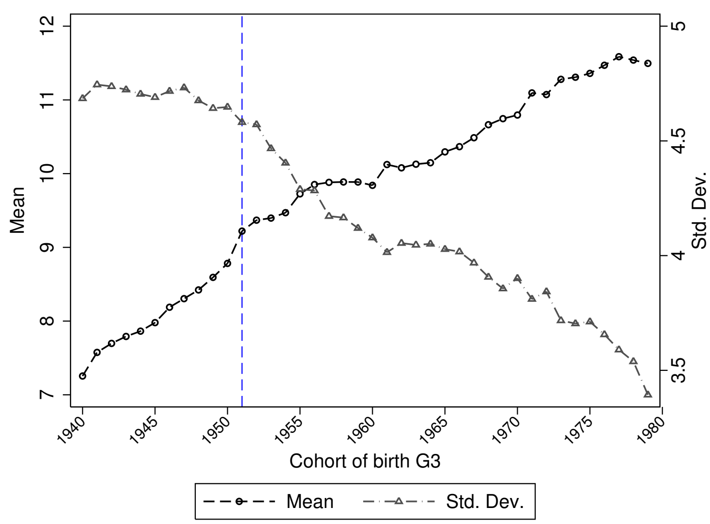
Notes: This figure shows the average standardized years of schooling of generation t in the y-axis, against the years of schooling of generation t1 in the x-axis. The triangle dots correspond to the analysis of G1 and G2, while the circular dots represent the analysis of G3 and G2. The diagonal is a 45-degree line.

Figure 4.5: Trends in intergenerational schooling coefficients



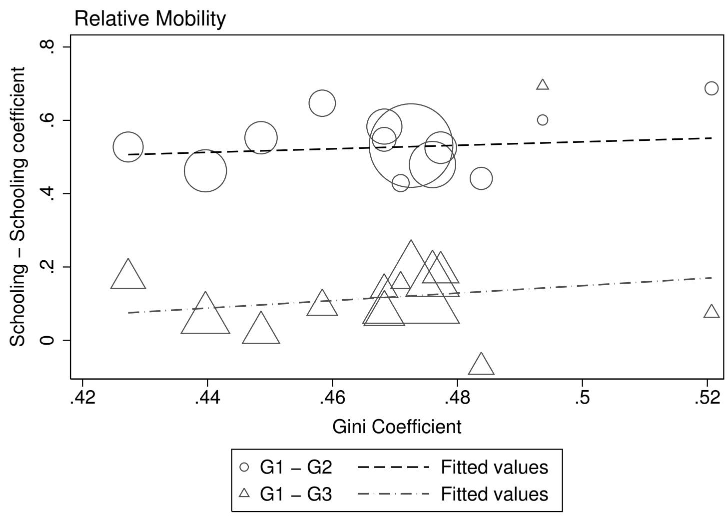
Notes: This figure shows in the y-axis the coefficients between a regression of G3 years of schooling against G2 years of schooling (in circles), and against G1 years of schooling (in triangles), for each birth cohort in Table 4.3.

Figure 4.6: Mean and SD of years of schooling by cohort of birth of G3



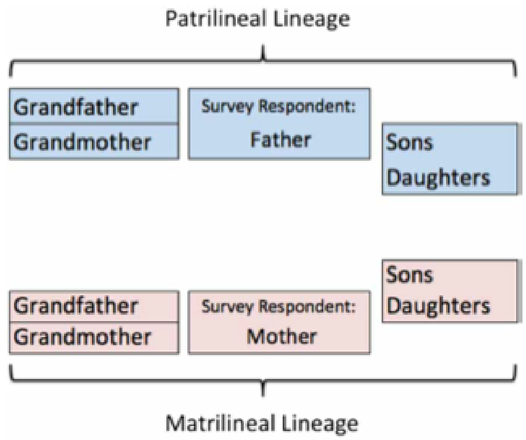
Notes: This figure shows in the left y-axis the average years of schooling and in the right y-axis the SD of years of schooling for cohorts born between 1940 and 1980. Numbers are computed using the Census of year 2002.

Figure 4.7: Correlation between Gini coefficient and intergenerational schooling coefficients for regions in Chile (Gatsby Curve)



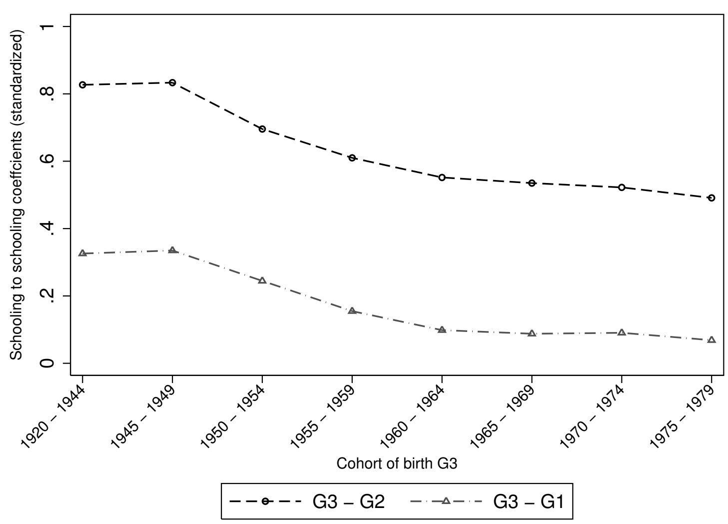
Notes: The left panel plots in the y-axis the coefficients between a regression of G2 years of education against G1 years of education (in circles) and a regression between G3 years of schooling against G1 years of schooling (in triangles), for each region of the country against the median Gini coefficient, in the x-axis, of the region as shown in Agostini and Brown (2007). The right panel plots the similar figures but using the coefficients of a regression of G2 rate of higher education attendance against G1 years of education (in circles) and a regression between G3 rate of higher education attendance and G1 years of education (in triangles).

Figure 4.8: Gender structure of the data



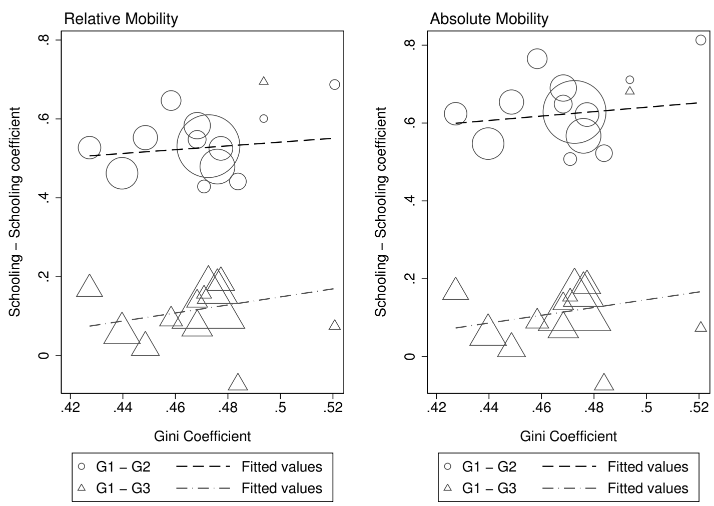
Notes: This figure describes the gender structure of our survey data, which allow to describe patrilineal and matrilineal lineages.Regarding the patrilineal lineage: If the survey respondent was the father of the household, we have information on the paternal grandparents and children. Regarding the matrilineal lineage: If the survey respondent was the mother of the household, we have information on the maternal grandparents and children.

Figure 4.9: Trends in intergenerational standardized schooling coefficients



Notes: This figure shows in the y-axis the coefficients between a regression of G3 standardized years of schooling against G2 standardized years of schooling (in circles), and against G1 standardized years of schooling (in triangles), for each birth cohort in Table 4.3.

Figure 4.10: Correlation between Gini coefficient and Intergenerational standardized schooling coefficients for regions in Chile (Gatsby Curve)



Notes: The left panel plots in the y-axis the coefficients between a regression of G2 standardized years of schooling against G1 standardized years of schooling (in circles) and a regression between G3 standardized years of schooling against G1 standardized years of schooling (in triangles), for each region of the country against the median Gini coefficient, in the x-axis, of the region as shown in Agostini and Brown (2007). The right panel plots the similar figures but using the coefficients of a regression using years of schooling without standardization.

Table 4.1: Descriptive statistics of the sample

Generation	Variable	Obs.	Mean	S.D.	Min	Max
Grandparents (G1)	G1 Schooling	4,529	4.38	4.08	0	19
	G1 Higher Education (=1)	4,529	0.03	0.17	0	1
	Grandfather Schooling	3,778	5.13	4.66	0	20
	Grandmother Schooling	4,233	4.05	4.03	0	19
	Grandfather Alive	4,219	0.17	0.38	0	1
	Grandmother Alive	4,509	0.33	0.47	0	1
	Any in G1 alive	4,529	0.38	0.49	0	1
Parents (G2)	G2 Schooling	4,529	7.99	4.72	0	20
	G2 Higher Education (=1)	4,529	0.15	0.36	0	1
	G2 Age	4,529	59.75	11.47	38	99
	G2 Male (=1)	4,529	0.52	0.5	0	1
	G2 Prim. Sch. in M.R. (=1)	4,529	0.29	0.45	0	1
	G2 Schooling (Father)	2,134	7.92	4.67	0	20
	G2 Schooling (Mother)	2,395	8.05	4.77	0	20
Children (G3)	G3 Schooling	12648	11.34	3.75	0	20
	G3 Higher Education (=1)	12648	0.27	0.44	0	1
	G3 Age	12648	35.65	9.17	23	82
	G3 Male (=1)	12648	0.51	0.5	0	1
	G3 Schooling (Daughter)	6219	11.36	3.66	1	20
	G3 Schooling (Son)	6429	11.32	3.83	0	20

Notes: All statistics are computed using the household sampling weights provided by the LSPS 2002. In total, we have 4529 individuals interviewed (G2) that provide information on their parents (G1) and all of their offspring (G3). G1 Schooling refers to the average years of education between the respondent' father and mother, who are the grandparents in our sample. All the rest of the "Schooling variables refer to the average years of completed education for the corresponding group. The Higher Education binary indicators are equal to 1 if years of schooling are above 12 years and equal to 0 if not. The variable G2 Prim. Sch. in MR is equal to one if G2 attended primary school in the MR of Chile, and equal to zero otherwise.

Table 4.2: Schooling coefficients: baseline results

Dep. Var. is Schooling of:	(1) G2	(2) G3	(3) G3	(4) G3
Schooling of G2		0.462***		0.400***
		(0.010)		(0.013)
		[0.575]		[0.499]
Schooling of G1	0.679***		0.388***	0.113***
	(0.016)		(0.015)	(0.015)
	[0.586]		[0.413]	[0.120]
Observations	4,529	12,648	12,648	12,648
Adjusted R2	0.423	0.369	0.234	0.378

Notes: Robust standard errors in parentheses clustered at the family level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions are estimated by ordinary least squares (OLS). Standardized coefficients are given within parentheses. G1: Mean schooling of the grandparent generation. G2: Schooling of the survey respondent. G3: Schooling of respondent's offspring. Column (1): Controls for a quadratic polynomial on age, sex of G2 and a binary indicator for whether G1 is alive. Each observation corresponds to a respondent in our sample. Column (2): Controls for a quadratic polynomial on age and sex of G3 and for whether G1 is alive. Column (3): Controls for a quadratic polynomial on age and sex of G2 and G3. Each observation corresponds to a child of the respondent. Column (4): Adds control for whether G1 is alive. Each term between brackets ([]) corresponds to the standardized coefficients.

Table 4.3: Distribution of birth cohorts of the children generation

Cohort group	Obs.	Percentage
1920-1944	244	1.90%
1945-1949	391	3.20%
1950-1954	817	6.70%
1955-1959	1,312	10.50%
1960-1964	2,064	16.30%
1965-1969	2,236	17.80%
1970-1974	2,728	21.40%
1975-1979	2,856	22.20%
Total	12,648	100.00%

Notes: This table shows the distribution of the offspring generation (G3) by five-year birth cohort groups. The first column shows the sample size in each birth cohort and the second column shows how birth cohorts of G3 are distributed in the sample.

Table 4.4: Cohort differences in intergenerational coefficients: G1-G3 and G2-G3

Dep. Var. is Schooling of:	(1) G3	(2) G3
Schooling of G1		0.320***
		(0.095)
Schooling of G2	0.685***	0.401***
	(0.053)	(0.013)
Schooling (G2) x Chrt: 1945 - 1949	0.005	
	(0.077)	
Schooling (G2) x Chrt: 1950 - 1954	-0.109	
	(0.070)	
Schooling (G2) x Chrt: 1955 - 1959	-0.180***	
	(0.059)	
Schooling (G2) x Chrt: 1960 - 1964	-0.228***	
	(0.056)	
Schooling (G2) x Chrt: 1965 - 1969	-0.242***	
	(0.056)	
Schooling (G2) x Chrt: 1970 - 1974	-0.252***	
	(0.055)	
Schooling (G2) x Chrt: 1975 - 1979	-0.278***	
	(0.054)	
Schooling (G1) x Chrt: 1945 - 1949		0.009
		(0.113)
Schooling (G1) x Chrt: 1950 - 1954		-0.08
		(0.094)
Schooling (G1) x Chrt: 1955 - 1959		-0.168*
		(0.099)
Schooling (G1) x Chrt: 1960 - 1964		-0.223**
		(0.095)
Schooling (G1) x Chrt: 1965 - 1969		-0.234**
		(0.096)
Schooling (G1) x Chrt: 1970 - 1974		-0.231**
		(0.095)
Schooling (G1) x Chrt: 1975 - 1979		-0.252***
		(0.095)
Adjusted R2	0.375	0.382
Observations	12,648	12,648
F-test on interactions	287	8.678

Notes: Robust standard errors in parentheses clustered at the family level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions are estimated by OLS, where the dependent variable is years of schooling of G3. The base group for birth cohorts is 19201944, the omitted variable. All coefficients are in deviations from this group. We control for main effects in all regressions. Column (1): Controls for a quadratic polynomial on age and a binary indicator for sex of G2 and G3. Column (2): Adds control for whether G1 is alive. Main effects and additional controls are omitted from the table for presentation purposes.

Table 4.5: Regional variations

Dep. Var. is Schooling of:	(1) G2	(2) G3
Schooling of G1	0.629***	0.105***
	(0.026)	(0.015)
Schooling (G1) x Region 1	-0.124	-0.217***
	(0.097)	(0.063)
Schooling (G1) x Region 2	-0.018	0.019
	(0.104)	(0.044)
Schooling (G1) x Region 3	-0.134	-0.007
	(0.082)	(0.053)
Schooling (G1) x Region 4	0.120*	-0.059*
	(0.062)	(0.033)
Schooling (G1) x Region 5	-0.074	-0.028
	(0.047)	(0.026)
Schooling (G1) x Region 6	-0.003	0.097***
	(0.070)	(0.034)
Schooling (G1) x Region 7	0.02	-0.041
	(0.071)	(0.029)
Schooling (G1) x Region 8	-0.051	0.057**
	(0.046)	(0.023)
Schooling (G1) x Region 9	-0.027	0.057**
	(0.072)	(0.029)
Schooling (G1) x Region 10	0.064	0.044
	(0.064)	(0.032)
Schooling (G1) x Region 11	0.146	0.327*
	(0.226)	(0.181)
Schooling (G1) x Region 12	0.184	-0.038
	(0.148)	(0.070)
Schooling (G1) x Region 13	-0.109	0.102***
	(0.174)	(0.033)
Adjusted R2	0.475	0.383
Observations	4,529	12,648
F-test of interactions	120.17	17.21

Notes: Robust standard errors in parentheses clustered at the family level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions are estimated by OLS where the dependent variable is years of schooling of G2 and years of schooling of G3, respectively. The base group for the regional interactions is the MR, the omitted variable. All coefficients are in deviations from this region. We control for main effects in all regressions. Column (1): Controls for a quadratic polynomial on age and a binary indicator for sex of G2 and G3. Column (2): Adds control for whether G1 is alive. Main effects and additional controls are omitted from the table for presentation purposes.

Table 4.6: Patrilineal schooling coefficients

Dep. Var. is Schooling of:	(1) Son	(2) Son	(3) Daughter	(4) Daughter
Father	0.436***	0.416***	0.402***	0.362***
	(0.020)	(0.022)	(0.020)	(0.021)
Paternal Grandfather	0.09***	0.059**	0.085***	0.013
	(0.022)	(0.025)	(0.021)	(0.026)
Paternal Grandmother		0.077**		0.159***
		(0.031)		(0.032)
Adjusted R2	0.408	0.41	0.392	0.403
Observations	2577	2577	2414	2414

Notes: Robust standard errors in parentheses clustered at the family level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions are estimated by OLS and control for age and age squared of sons/daughters and fathers. From the 4529 respondents, 2395 are men 2134 are women. Following the patrilineal lineage, the male respondents report to have an offspring of 7029, divided in 3647 sons and 3382 daughters. The number of sons with complete information on their grandmother's and grandfather's schooling, parent education, and age is of 2577. The same number for the daughters is 2414.

Table 4.7: Matrilineal schooling coefficients

Dep. Var. is Schooling of:	(1) Daughter	(2) Daughter	(3) Son	(4) Son
Mother	0.426***	0.413***	0.422***	0.395***
	(0.023)	(0.024)	(0.028)	(0.029)
Maternal Grandmother	0.079***	0.053**	0.105***	0.042
	(0.028)	(0.026)	(0.031)	(0.036)
Maternal Grandfather		0.047		0.102***
		(0.030)		(0.030)
Adjusted R2	0.414	0.416	0.393	0.4
Observations	2033	2033	2033	2033

Notes: Robust standard errors in parentheses clustered at the family level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions are estimated by OLS and control for age and age squared of sons/daughters and parents. From the 4529 respondents, 2395 are men 2134 are women. Regarding the matrilineal lineage, the 2134 women report having 5698 children older than 23 years of age. Of them, 2827 are sons and 2871 are daughters. The number of them with complete information is 2333 for sons and daughters.

Table 4.8: Full models and IPW corrections

Dep. Var. is Schooling of:	(1) G2	(2) G3	(3) G3	(4) G3	(5) G2 (IPW)	(6) G3 (IPW)	(7) G3 (IPW)	(8) G3 (IPW)
Schooling of G2		0.462***		0.400***		0.464***		0.404***
		(0.010)		(0.013)		(0.010)		(0.013)
Schooling of G1	0.679***		0.388***	0.113***	0.670***		0.390***	0.113***
	(0.016)		(0.015)	(0.015)	(0.016)		(0.014)	(0.014)
Age of G2	-0.078	0.244***		0.237***	-0.005	0.250***		0.241***
	(0.052)	(0.045)		(0.044)	(0.049)	(0.044)		(0.043)
Age of G2 square	0	-0.002***		-0.002***	-0.001	-0.002***		-0.002***
	0.000	0.000		0.000	0.000	0.000		0.000
Sex of G2 (Male=1)	0.434***	-0.353***	-0.039	-0.297***	0.346***	-0.369***	-0.118	-0.323***
	(0.119)	(0.094)	(0.109)	(0.093)	(0.119)	(0.092)	(0.106)	(0.091)
G1 still lives (=1)	0.326**		0.058	0.029	0.361**		0.026	-0.022
	(0.147)		(0.115)	(0.101)	(0.148)		(0.113)	(0.102)
Age of G3		0.026	0.075**	0.025		0.038	0.082**	0.035
		(0.033)	(0.032)	(0.032)		(0.032)	(0.033)	(0.031)
Age of G3 square		-0.001***	-0.002***	-0.001***		-0.001***	-0.002***	-0.001***
		0.000	0.000	0.000		0.000	0.000	0.000
Sex of G3 (Male=1)		-0.138**	-0.061	-0.127**		-0.129**	-0.053	-0.118**
		(0.056)	(0.061)	(0.055)		(0.056)	(0.062)	(0.055)
Constant	9.354***	0.473	9.941***	0.673	7.165***	0.114	9.845***	0.483
	(1.602)	(1.265)	(0.608)	(1.244)	(1.517)	(1.186)	(0.606)	(1.190)
Observations	4,529	12,648	12,648	12,648	4,529	12,648	12,648	12,648
Adjusted R2	0.423	0.369	0.234	0.378	0.409	0.372	0.233	0.381

Notes: Robust standard errors in parenthesis clustered at the family level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. G1: Mean schooling of the grandparent generation. G2: Schooling of the survey respondent. G3: Schooling of respondent's offspring. Column (1): OLS of G2 Schooling on G1 Schooling, and controls for a quadratic polynomial on age and a binary indicator of sex of G2 and a binary indicator for whether any grandparent is still alive. Each observation corresponds to a respondent in our sample. Column (2): OLS of G3 Schooling on G2 Schooling, and controls for a quadratic polynomial on age and a binary indicator of sex of G2 and a quadratic polynomial of age and a binary indicator of sex of G3. Each observation corresponds to a child of the respondent. Column (3): OLS of G3 Schooling on G2 and G1 Schooling, and controls for a quadratic polynomial on age and a binary indicator of sex of G2, a quadratic polynomial of age and a binary indicator of sex of G3, and a binary indicator for whether any grandparent is still alive.

Table 4.9: Regional variations: full tables

Dep. Var. is:	(1) Sch. G2	(2) H.Ed. G2	(3) Sch. G3	(4) H.Ed. G3
Schooling of G1	0.629***	0.032***	0.105***	0.011***
	(0.026)	(0.002)	(0.015)	(0.002)
Schooling (G1) x Region 1	-0.124	0	-0.217***	-0.005
	(0.097)	(0.010)	(0.063)	(0.007)
Schooling (G1) x Region 2	-0.018	-0.007	0.019	0.002
	(0.104)	(0.007)	(0.044)	(0.006)
Schooling (G1) x Region 3	-0.134	0.018	-0.007	-0.001
	(0.082)	(0.020)	(0.053)	(0.006)
Schooling (G1) x Region 4	0.120*	0.001	-0.059*	-0.004
	(0.062)	(0.006)	(0.033)	(0.004)
Schooling (G1) x Region 5	-0.074	-0.004	-0.028	-0.006*
	(0.047)	(0.004)	(0.026)	(0.003)
Schooling (G1) x Region 6	-0.003	-0.004	0.097***	0.003
	(0.070)	(0.007)	(0.034)	(0.004)
Schooling (G1) x Region 7	0.02	-0.002	-0.041	-0.004
	(0.071)	(0.006)	(0.029)	(0.004)
Schooling (G1) x Region 8	-0.051	-0.009**	0.057**	-0.002
	(0.046)	(0.004)	(0.023)	(0.003)
Schooling (G1) x Region 9	-0.027	-0.006	0.057**	0
	(0.072)	(0.005)	(0.029)	(0.004)
Schooling (G1) x Region 10	0.064	0	0.044	-0.006
	(0.064)	(0.006)	(0.032)	(0.004)
Schooling (G1) x Region 11	0.146	0.009	0.327*	0.022
	(0.226)	(0.019)	(0.181)	(0.022)
Schooling (G1) x Region 12	0.184	0.035	-0.038	0.017
	(0.148)	(0.025)	(0.070)	(0.017)
Schooling (G1) x Region 13	-0.109	-0.014	0.102***	0.002
	(0.174)	(0.009)	(0.033)	(0.005)
Age of G2	-0.133***	0.012**	0.229***	0.024***
	(0.050)	(0.006)	(0.029)	(0.004)
Age of G2 square	0.001	-0.000**	-0.001***	-0.000***
	0.000	0.000	0.000	0.000
Sex of G2 (Male=1)	0.447***	0.002	-0.284***	-0.040***
	(0.115)	(0.011)	(0.058)	(0.007)
Region 1	1.275	-0.005	1.397***	0.028
	(0.792)	(0.088)	(0.368)	(0.049)
Region 2	0.703	0.063	0.403	0.046
	(0.612)	(0.054)	(0.328)	(0.046)
Region 3	0.218	-0.315*	0.257	0.015
	(0.538)	(0.187)	(0.383)	(0.045)
Region 4	-0.942***	-0.019	0.669***	0.076***
	(0.313)	(0.044)	(0.182)	(0.024)
Region 5	-0.067	-0.015	0.334**	0.052**
	(0.276)	(0.035)	(0.169)	(0.021)

Table 4.9 (continued)

Dep. Var. is:	(1) Sch. G2	(2) H.Ed. G2	(3) Sch. G3	(4) H.Ed. G3
Region 6	-0.672**	0.007	0.03	0.026
	(0.327)	(0.044)	(0.170)	(0.022)
Region 7	-0.627**	0.008	-0.042	0.003
	(0.303)	(0.035)	(0.154)	(0.021)
Region 8	-0.458*	0.027	-0.22	0.032*
	(0.259)	(0.033)	(0.137)	(0.017)
Region 9	-0.267	0.047	-0.082	0
	(0.306)	(0.034)	(0.157)	(0.021)
Region 10	-1.268***	-0.033	-0.601***	0.036
	(0.270)	(0.039)	(0.174)	(0.022)
Region 11	-0.648	-0.129	-1.308	-0.007
	(1.029)	(0.079)	(0.800)	(0.088)
Region 12	0.059	-0.332	0.826	0.011
	(1.021)	(0.233)	(0.616)	(0.097)
Region 13	-4.788***	0.026	-0.380**	0.044*
	(0.334)	(0.056)	(0.151)	(0.024)
G1 still lives (=1)	0.312**	0.023*	0.039	0.017**
	(0.142)	(0.012)	(0.067)	(0.009)
Age of G3			0.025	-0.013***
			(0.025)	(0.003)
Age of G3 square			-0.001***	0.000**
			0.000	0.000
Sex of G3 (Male=1)			-0.122**	-0.001
			(0.055)	(0.007)
Schooling of G2			0.391***	0.038***
			(0.008)	(0.001)
Constant	11.348***		0.931	
	(1.540)		(0.775)	
Adjusted R2	0.475	0.232	0.383	0.276
Observations	4,529	4,529	12,648	12,648

Notes: Robust standard errors in parentheses clustered at the family level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions are estimated by OLS. Column (1) uses years of schooling of G2 as the dependent variable, column (2) uses a binary variable for whether G2 attended higher education or not, column (3) uses years of schooling of G3 as the dependent variable, and column (4) uses a binary variable for whether G3 attended higher education or not. Columns (2) and (4) report marginal effects of a probit model. The R2 for these two columns corresponds to the pseudo-R2 from the probit models.

Table 4.10: Differences by region in census 1960

						Region of Chile							
	I	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**	**XII**	**R.M.**
Demographics (%)													
Less than primary Sch.	0.62	0.60	0.65	0.79	0.65	0.80	0.84	0.78	0.81	0.83	0.85	0.65	0.61
Not connected sewage	0.48	0.42	0.61	0.72	0.37	0.75	0.72	0.69	0.84	0.82	0.89	0.48	0.33
No piped water	0.44	0.29	0.54	0.66	0.30	0.69	0.63	0.54	0.72	0.76	0.66	0.35	0.19
Industries (%)													
Agriculture and others	0.13	0.02	0.20	0.42	0.23	0.55	0.59	0.38	0.56	0.51	0.42	0.17	0.09
Mining	0.08	0.36	0.24	0.06	0.01	0.06	0.00	0.06	0.00	0.00	0.01	0.09	0.00
Manufacturing	0.16	0.10	0.08	0.10	0.18	0.08	0.08	0.16	0.08	0.12	0.15	0.14	0.26
Electricity, gas and water	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.04	0.00	0.01
Construction	0.08	0.10	0.12	0.06	0.07	0.05	0.05	0.07	0.04	0.06	0.03	0.08	0.09
Wholesale and retail trade	0.13	0.09	0.08	0.10	0.11	0.09	0.07	0.08	0.07	0.07	0.08	0.09	0.15
Hotels and restaurants	0.02	0.02	0.01	0.02	0.03	0.01	0.01	0.01	0.01	0.01	0.02	0.03	0.03
Transportation and comm.	0.10	0.11	0.10	0.06	0.08	0.03	0.03	0.05	0.04	0.05	0.05	0.05	0.06
Financial services	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.02
Public administration	0.18	0.05	0.03	0.02	0.09	0.02	0.03	0.05	0.05	0.03	0.09	0.21	0.06
Real estate	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Education	0.02	0.04	0.04	0.04	0.04	0.02	0.02	0.03	0.03	0.02	0.01	0.02	0.03
Health and social work	0.01	0.02	0.01	0.01	0.03	0.01	0.02	0.01	0.01	0.01	0.01	0.02	0.04
Other services	0.03	0.03	0.04	0.02	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.04
Private household services	0.04	0.05	0.05	0.08	0.08	0.06	0.07	0.08	0.09	0.08	0.07	0.06	0.12

Notes: All statistics are computed from the Chilean Census 1960 sample provided in IPUMS-International. The percentage of people with less than primary school is computed among those born between 1939 and 1941, which corresponds, to the median cohort of the parental generation in our sample. The percentage of households not connected to sewage system and the percentage of households not connected to pied water are computed at the household level using household weights provided by the Census. The percentage of people within industries is computed among all those observations that are employed. The industry codes correspond to International Industry Codes.

Table 4.11: Patrilineal schooling standardized coefficients

Dep. Var. is Schooling of:	(1) Son	(2) Son	(3) Daughter	(4) Daughter
Father	0.530***	0.505***	0.489***	0.440***
	(0.024)	(0.026)	(0.024)	(0.026)
Paternal Grandfather	0.106***	0.070**	0.101***	0.015
	(0.026)	(0.030)	(0.025)	(0.031)
Paternal Grandmother		0.076**		0.158***
		(0.031)		(0.032)
Adjusted R2	0.408	0.41	0.392	0.403
Observations	2577	2577	2414	2414

Notes: Robust standard errors in parentheses clustered at the family level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions are estimated by OLS and control for age and age squared of sons/daughters and fathers. From the 4529 respondents, 2395 are men 2134 are women. Following the patrilineal lineage, the male respondents report to have an offspring of 7029, divided in 3647 sons and 3382 daughters. The number of sons with complete information on their grandmother's and grandfather's schooling, parent education, and age is of 2577. The same number for the daughters is 2414.

Table 4.12: Matrilineal schooling standardized coefficients

Dep. Var. is Schooling of:	(1) Daughter	(2) Daughter	(3) Son	(4) Son
Mother	0.486***	0.470***	0.481***	0.450***
	(0.026)	(0.028)	(0.032)	(0.033)
Maternal Grandmother	0.079***	0.063**	0.104***	0.042
	(0.028)	(0.031)	(0.031)	(0.036)
Maternal Grandfather		0.046		0.121***
		(0.029)		(0.036)
Adjusted R2	0.414	0.416	0.393	0.4
Observations	2033	2033	2033	2033

Notes: Robust standard errors in parentheses clustered at the family level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions are estimated by OLS and control for age and age squared of sons/daughters and parents. From the 4529 respondents, 2395 are men 2134 are women. Regarding the matrilineal lineage, the 2134 women report having 5698 children older than 23 years of age. Of them, 2827 are sons and 2871 are daughters. The number of them with complete information is 2333 for sons and daughters.