ISSN 2279-9362

Parental Love Is Not Blind: Identifying Selection into Early School Start

Ainoa Aparicio Fenoll, Nadia Campaniello and Ignacio Monzon

No. 704
December 2023

Carlo Alberto Notebooks

www.carloalberto.org/research/working-papers

© 2023 by Ainoa Aparicio Fenoll, Nadia Campaniello and Ignacio Monzon. Any opinions expressed here are those of the authors and not those of the Collegio Carlo Alberto.

Parental Love Is Not Blind: Identifying Selection into Early School Start*

Ainoa Aparicio Fenoll† Nadia Campaniello‡ Ignacio Monzón§

October 24, 2023

Abstract

Do parents take into account their children's ability when deciding on their education? If so, are parents' perceptions accurate? We study this by analyzing a key educational decision. Parents choose whether their children start elementary school one year early. Do they select high ability kids to start early? We propose a novel methodology to identify the sign and strength of selection into early starting. We find robust evidence of positive selection. Had they started regularly, early starters would have obtained test scores 0.2 standard deviations higher than the average student. Our simple methodology applies to RDD settings in general.

JEL Classification: I24; C21; J13

Keywords: School starting age; Selection; Children; Education; Treatment effects

*We are grateful to Sule Alan, Orazio Attanasio, Philip Cook, Rajeev Dehejia, Emilia Del Bono, Adeline Delavande, Raquel Fernández, Libertad González Luna, Eric Hanushek, Toomas Hinnosaar, Joseph Hotz, Melanie Luhrmann, Giovanni Mastrobuoni, Juan S. Morales, Matthias Parey, Daniele Paserman, Birgitta Rabe, Jeff Smith, Alex Tetenov and Sarid Weisburd for their helpful comments. We thank seminars participants at Collegio Carlo Alberto, IMT Lucca, the University of Padova and the University of Essex, as well as participants at the Alpine Population Conference, the 2018 Meetings of the SEHO, the Comparative Law and Economics Forum Meeting, the 2018 EALE Conference, the 2018 Petralia Applied Economics Workshop, the 2018 Workshop on Perceptions and Public Policies, the First Unibg Winter Symposium in Economics and the Third Bolzano Applied Microeconomics Workshop.

†University of Turin (ESOMAS Department) and Collegio Carlo Alberto; ainoa.apariciofenoll@unito.it

‡University of Turin (ESOMAS Department) and Collegio Carlo Alberto; nadia.campaniello@unito.it

§Collegio Carlo Alberto and University of Turin (ESOMAS Department); ignacio@carloalberto.org

1 Introduction

Parents make many crucial decisions regarding the education of their children. They decide whether to enroll their children in early formal child care, which school they attend, whether they participate in extracurricular activities, and their school starting age. These are critical decisions with effects that may depend on each child's characteristics. While a demanding school (or engaging in many extracurricular activities) may be beneficial for high-ability children, it may harm low-ability ones. Do parents take into account their children's characteristics when making these decisions?

We focus on parents' decisions on school starting age. In many countries, parents can choose that their children start elementary school one year early. The literature on early child development shows that this decision has lasting consequences. On the one hand, early starters enter the labor market one year early, which increases their returns to human capital (see for example Black, Devereux, and Salvanes [2011]). Moreover, children from disadvantaged backgrounds could benefit from attending school rather than staying home. On the other hand, early starters have worse academic performance. The relative importance of these factors and, hence, the effect of starting early can depend on each child's characteristics. Do parents take into account their children's characteristics when deciding whether they start early?

The main difficulty in understanding which students are selected into early starting is that early starting itself affects students' academic outcomes. To see this, consider children who differ in their underlying unobservable ability. We would like to know whether high-ability children are selected into early starting.[1] Everything else equal, a higher-ability student obtains higher test scores. However, age also affects scores: an extra month leads to higher test scores. To compare the underlying ability of those who start one year early relative to those who do not, a naïve approach would be to compare the test scores of these two groups. Unfortunately, the score difference depends on both (i) the difference in ability and (ii) the difference in age. Students who start one year early are effectively

[1]Ideally, one would use data on children's ability *before* the decision of early starting. However, the measures of cognitive ability for children of pre-school age are limited. We thus rely on (mandatory) standardized test scores during elementary school.

twelve months younger than regular students born in the same month. Being twelve months younger has a strong negative effect on test scores. Therefore, we could mistakenly conclude that early starters are low-ability students.

Our strategy to identify the characteristics of students who start early is based on (i) a feature of the Italian education system and (ii) an empirical regularity on how age affects test scores. First, children in Italy can start elementary school one year early *only if they are born between January and April*. Children born between May and December cannot choose to start one year early. Thus, children within the same class can be born up to sixteen months apart. Second, there is a pattern on how age affects academic performance: average test scores increase linearly with each extra month of age. Crawford, Dearden, and Meghir [2010] show this empirical regularity for England, while Black et al. [2011] do it for Norway, Fredriksson and Öckert [2014] for Sweden, and Cook and Kang [2018] for the US. Reassuringly, the linearity of test scores on month of birth holds for countries with different school cutoff dates. We document this also for our Italian dataset: the age-in-months effect on average test scores is linear for children born between May and December.

Our methodology to identify which students are selected into early starting has three steps. First, we use the subsample of children born between May and December to estimate age-in-months effects on test scores. Children born in these months cannot choose to start early. Hence, differences in average test scores across months of birth are exclusively due to age differences when taking the test. Second, we use the estimated age-in-months effects to compute the average test scores for *all* children born between January and April, had *all* of them started regularly. In practice, we extrapolate the linear trend found for children born between May and December to the months between January and April. Finally, we compute the average test scores that early starters would have obtained had they started regularly. If the average test scores that early starters would have obtained are higher than those of the average student in the population, we conclude that there is positive selection. Our measure of the strength of selection is the difference between the average test score of early starters, had they started regularly, and the average test score in the population, had all students started regularly.

We use data on standardized tests administered to all students in Italy. These tests cover two subjects (mathematics and Italian), are designed by an agency of the Italian Government (the National Institute for the Evaluation of the School System - INVALSI) and are mandatory for all students. Students are tested in the second, fifth, eighth and tenth grades of compulsory schooling. We use information on test scores, month and year of birth, and students' and parents' characteristics. Our sample covers academic years 2011–12 to 2018–19. For our main results, we use data from grade two, which is the closest to the decision to start early.

Our main result is that early starters are positively selected. Had they started regularly, early starters would have been at the top of the grade distribution of their cohort: they would have obtained (on average) scores 0.2 standard deviations higher than the average student. This pattern of positive selection arises for all cohorts and for all months of birth (January to April).

Our methodology to measure the strength of selection can also be applied to study questions other than selection into early starting. Our simple approach is valid whenever (i) a well-defined functional form relates the outcome and running variable and (ii) there is an exogenous cutoff in the running variable. As these conditions hold in most regression discontinuity design settings, our methodology can be applied to those contexts.

Our second main result quantifies the penalty from early starting. Early starters obtain lower test scores because they start elementary school twelve months ahead. We define the penalty from early starting as the magnitude of this decrease. A comparison between the test scores of early starters and those of regular starters would not reflect the penalty from early starting. Since early starters are positively selected, such a comparison would underestimate the penalty from early starting. By taking advantage of our measure for the strength of selection, we quantify the penalty both for selected students and for the average student in the population. We find that selected students born in January and February suffer a penalty as large as that of the average student in the population. Selected students born in March and April instead suffer a penalty *lower* than that of the average student in the population.

We present a set of extensions and robustness checks for our results. First, we show that our results hold even after controlling for a rich set of observable characteristics. Second, we test for regional differences across Italy and find that selection is stronger in Southern Italy. Third, we provide further evidence for the linear relationship between test scores and age in months. Finally, we apply our methodology to data from grade five instead of grade two and also find strong evidence of positive selection.

1.1 Related literature

A growing literature studies parents' perceptions about their children's abilities and how these perceptions shape parents' decisions. Kinsler and Pavan [2021] find that parents' beliefs about their children's skills relative to children of the same age are determined by their children's skills relative to children of the same school. They also find a positive relationship between children's perceived abilities and parents' investment in human capital. Dizon-Ross [2019] shows that parents, especially the poorer and less educated, have inaccurate beliefs about their children's performance at school. In turn, these inaccurate perceptions prevent parents from investing optimally in their children's human capital. Once provided with the correct information, parents invest more efficiently in their education.

We use an indirect approach to study parents' perceptions and how these perceptions shape parents' decisions. Unlike other papers in the literature, we do not measure parents' perceptions directly. Instead, we observe the decision of parents to have their children start elementary school one year early. We identify the strength of selection, which reflects the ability of selected children. We find that parents select higher-ability children to start one year early. Thus, our indirect approach sheds light on parents' perceptions of their children's abilities and how these perceptions map into decisions.

Early work published in education journals analyzes the impact of age-in-months differences within the classroom on educational outcomes. There is consensus that older students obtain higher scores (see Russell and Startup [1986], Borg and Falzon [1995], Sharp [1995], Thomas [1995], Massey, Elliott, and Ross [1996], Sharp and Hutchison [1997] and Alton and Massey [1998]). In line with this literature, we find that older students in the classroom do better: an additional month of age is associated with a test score 0.03 standard deviations higher.[2]

A birth date cutoff (typically January 1st or September 1st) determines the school starting age in most countries. Students born before that cutoff start school one year earlier than students born after it. Several papers study the impact of school starting age on school performance using regression discontinuity designs (see McEwan and Shapiro [2008], Dobkin and Ferreira [2010], Crawford, Dearden, and Greaves [2014], Cook and Kang [2016] and Dhuey, Figlio, Karbownik, and Roth [2019]). They find that starting school one year later increases students' test scores. In line with these papers, we find that starting school one year *early* carries a penalty in test scores.

In some countries, parents can decide when their children start school.[3] Parents may take into account their perceptions on their children's ability when deciding to have them start early (or late). Therefore, consistent estimates of the impact of school starting age on students' performance must account for selection. To do this, several papers use instrumental variables (see Angrist and Krueger [1991], Bedard and Dhuey [2006], Elder and Lubotsky [2009] and Fredriksson and Öckert [2014]). A child's age relative to the cutoff is a common instrument for the actual school starting age. These papers also find that starting one year later increases tests scores, which is consistent with our findings.

We depart from previous work in that our focus is on identifying the strength of selection into early starting. Previous work instead focuses on the effect of relative age within the classroom (or school starting age) on several measures of students' performance. In our main contribution, we provide an intuitive methodology to identify the strength of selection into early starting.[4] Our methodology also applies to environments other than

[2]Differences in performance across months of birth can be interpreted as causal estimates when students are *as if* randomly assigned to months of birth. This is not the case in some countries (see for example Fan, Liu, and Chen [2017] and Clarke, Oreffice, and Quintana-Domeque [2019]). In Appendix A.5 we show that observed characteristics are practically identical across months of birth in our data. Thus, differences in characteristics across months of birth do not significantly affect our estimates.

[3]Redshirting is the practice of postponing entrance into kindergarten of age-eligible children. This practice is common in the US. Academic redshirting occurs at the rate of about 9% per year, according to the National Center for Education Statistics (West, Meek, and Hurst [2000]).

[4]In a related paper, Black, Joo, LaLonde, Smith, and Taylor [2022] propose a different methodology to test for the presence of selection and to assess the strength of selection. They illustrate their methodology using three well-known applications, including the causal impact of children on maternal labor supply as in Angrist and Evans [1998]. Black et al. [2022]'s methodology applies to cases where the causal estimate of

early starting. Whenever the causal effect can be estimated using a regression discontinuity design, our methodology measures the strength of selection into treatment.

Our second contribution is closer to previous work in the literature. We estimate the penalty from early starting; that is, the effect of starting school one year early on test scores. A naïve estimate of the penalty would result from computing the difference between the test scores of early starters and those of regular starters. However, such an estimate would be biased, as it does not account for selection. We use our methodology to estimate the strength of selection to obtain unbiased estimates of the penalty from early starting. Therefore, instead of relying on regression discontinuity designs or instrumental variables to estimate the penalty, we use our simple methodology, that relies on sample averages. Our approach allows for the comparison of the penalty from early starting between those students actually selected to early start and the average student in the population.[5] Our approach also provides separate estimates of the penalty by month of birth.

2 Data and institutional framework

We use standardized test score data from the National Institute for the Evaluation of the School System (INVALSI). Education is compulsory in Italy between ages 6 and 16. The education system is divided into elementary school (five years), middle school (three years), and secondary school (five years). We provide further institutional details in Appendix A.1. Students take standardized tests in the second and fifth year of elementary school, then three years later in the third year of middle school, and finally two years later in the second year of secondary school. INVALSI provides data from all academic years between 2009–10 and 2021–22, except for 2019–20. Because of Covid, no standardized

interest is obtained either from an experiment, or with instrumental variables with a binary instrument, or with regression discontinuity designs. Our dataset is observational instead of experimental. Moreover, we do not use instruments in our paper. Finally, an RDD would only provide estimates of selection for dates close to the cutoff.

[5] In related work, Ordine, Rose, and Sposato [2018] take advantage of the same administrative data as we do in this paper. They use a reduced form regression discontinuity design to estimate how *having the choice* to start early affects test scores. We instead study the effect on test scores of actually *choosing* to start early.

testing occurred during the academic year 2019–20.

The INVALSI data contains test scores from two subjects (mathematics and Italian) and indicates the number of correct answers. We standardize scores by subject, academic year and grade to have zero mean and unit variance (as in Angrist, Battistin, and Vuri [2017]). The data set also includes students' characteristics (among them: gender and whether they are foreign-born) and parental characteristics (among them: whether they are foreign-born, their level of education and labor market status). We describe these characteristics in detail in Appendix A.4.

We make a series of exclusions to arrive to the sample we use for our analysis. First, our sample only includes academic years 2011–12 to 2018–19. Information on students' month of birth is not available for academic years 2009–10 and 2010–11. We thus exclude them, as information on students' month of birth is crucial to identify selection. Since there is no data from academic year 2019–20, we also exclude academic years after 2018–19. We do this since we need data from two consecutive academic years to study selection. Our sample allows us to present results for students born between 2005 and 2011. All students born in the same year belong to the same cohort. Thus, we present results for seven cohorts.[6]

Second, our sample only includes grade two. Selection into early starting takes place right before the first year of elementary school. Since our objective is to identify selection, we focus on grade two, the closest to this decision. Moreover, the effects of an extra month of age on scores are stronger in grade two than in later grades. Finally, early starters may appear as regular starters if they repeat a grade. Grade repetition is highly uncommon in second grade. See Appendix A.3 for a discussion on using data from grade five.

Next, we include in our sample only regular and early starters. We say children are *regular* when they turn seven the year they start grade two (the standard age according to Italian law). Instead, children are *early starters* if they satisfy two conditions: they turn six the year they start grade two and they are born between January and April. We then exclude students from three groups. First, those who turn eight or more the year they

[6]We present results for consecutive cohorts. We could also present results for cohort 2014 (using data from academic years 2020–21 and 2021–22). We omit them since they are similar to those we present here.

start grade two (1.56% of total students in grade two). Second, those who turn five or less the year they start grade two (less than 0.01% of total students in grade two). Third, we also exclude students who turn six the year they start grade two but are born between May and December (0.37% of total students in grade two).

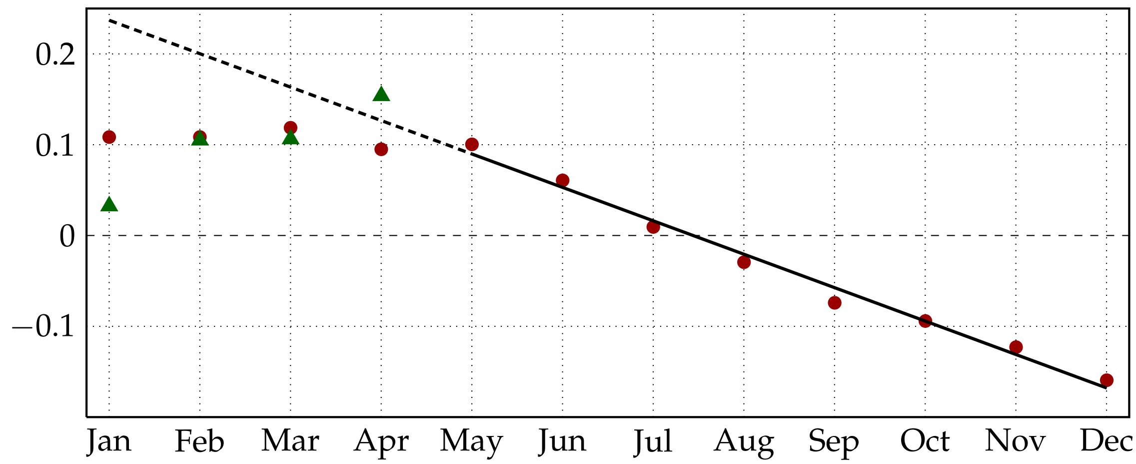
The resulting data set includes $3,287,893$ observations for the mathematics test and $3,266,288$ observations for the Italian test. In our sample, 31.9% of children are potential early starters since they are born between January and April. Of those, only 26.0% start early, so 8.27% of the observations in our sample correspond to early starters. Of all students born in January, 42.4% of them are early starters. This proportion decreases to 27.9% for February, 19.1% for March, and finally 13.3% for April.

Early starters are more likely to be female (54.8% instead of 47.1% for regular starters born in the same months), less likely to be foreign-born (1.1% instead of 2.3%), and less likely to have foreign-born parents (8.9% instead of 13.5%). They have a higher proportion of parents with university degrees (24.3% instead of 18%). We present descriptive statistics of students and their parents in Table 8 in Appendix A.4.[7]

Average test scores exhibit some common patterns for all cohorts and for both subjects. To illustrate these patterns, consider the test scores in mathematics for students born in 2011. Figure 1 presents average test scores by month of birth both for regular and early starters. Circles (in red) represent average test scores for regular starters, who enroll in second grade in academic year 2018–19. Triangles (in green) represent average test scores for early starters, who enroll in second grade in the previous academic year (2017–18). The thick line fits the average test scores of regular starters born between May and December. Average test scores exhibit a linear decrease from May to December. However, average test scores for *regular starters* born between January and April lie below this linear trend. This underperformance of regular starters is preliminary evidence of positive selection into early start.

[7]The methodology that we propose in the current paper measures the strength of selection, including selection on observables and unobservables. Section 5.1 discusses whether observable characteristics can explain selection. We find that there is positive and significant selection for all cohorts, even controlling for observable characteristics.

Figure 1: Average test scores of regular and early starters. Mathematics. Cohort 2011



Notes: Circles (in red) represent average test scores for regular starters, who enroll in second grade in academic year 2018–19. Triangles (in green) represent average test scores for early starters, who enroll in second grade in the previous academic year (2017–18). The thick line fits average test scores of regular starters born between May and December. The dashed line depicts its extrapolation to the months between January and April.

3 Methodology

Our first objective is to identify which students are selected into early starting: do higher ability students start early? In other words, had they started regularly, would selected students have obtained grades from the top of the distribution? Unfortunately, we do not observe their counterfactual scores. Moreover, the scores we observe from these students include a strong age effect: selected students are twelve months younger than non-selected students born in the same month.

We compute counterfactual average scores that account for the age effect. To do so, we express test scores $T^t(m, x)$ as a function of m (age-in-months) and x (any other individual characteristics that determine scores). The superscript t indexes academic years. A student with characteristics (m, x) who starts regularly obtains scores $T^t(m, x)$, while one who starts early obtains scores $T^{t-1}(m - 12, x)$.

Students may belong to one of three groups $G \in \{S, NS, U\}$, where S denotes students who are selected into early starting, NS denotes students who are not selected into early starting (that is, they start regularly) and $U = S \cup NS$ denotes all students. To construct our counterfactual scores, we consider two possible treatments $D \in \{E, R\}$, where E

denotes early starting, and R indicates starting regularly. Then, expected average scores for different groups are given by:

$$A^t(G, D, m) = \begin{cases} E\left[T^t(m, x_i) \mid i \in G, m\right] & \text{if } D = R \\ E\left[T^{t-1}(m-12, x_i) \mid i \in G, m\right] & \text{if } D = E \end{cases}$$

The strength of selection is given by $A(S, R, m) - A(U, R, m)$: the difference between the average test score of early starters, had they started regularly, and the average test score in the population, had all students started regularly.[8] Although we do not observe these magnitudes, we can indirectly infer them. Our methodology follows three steps.

Estimating age-in-months effects. Our methodology relies on the key identifying assumption that $A(U, R, m) = \alpha + \beta m$, that is, average test scores in the population are linear in age-in-months. As discussed in the introduction, there is evidence that this is the case for many countries with different school starting age cutoffs. We provide further evidence that average scores are linear in age-in-months also in our data in Appendix A.2.

In our first step, we estimate the linear age-in-months effect on test scores on the subsample of regular students born between May and December using the following equation:

$$T_i^{st} = \alpha^{st} + \beta^{st} m_i^t + \varepsilon_i^{st} \qquad \forall\ s, t, \text{ and for } m_i \in \{5, \ldots, 12\} \tag{1}$$

where T_i^{st} is the standardized test score in subject s and academic year t of student i born in month m_i. Our coefficient of interest β measures the effect of an extra month of age on test scores. We estimate equation (1) separately for each subject s and for each academic year t.

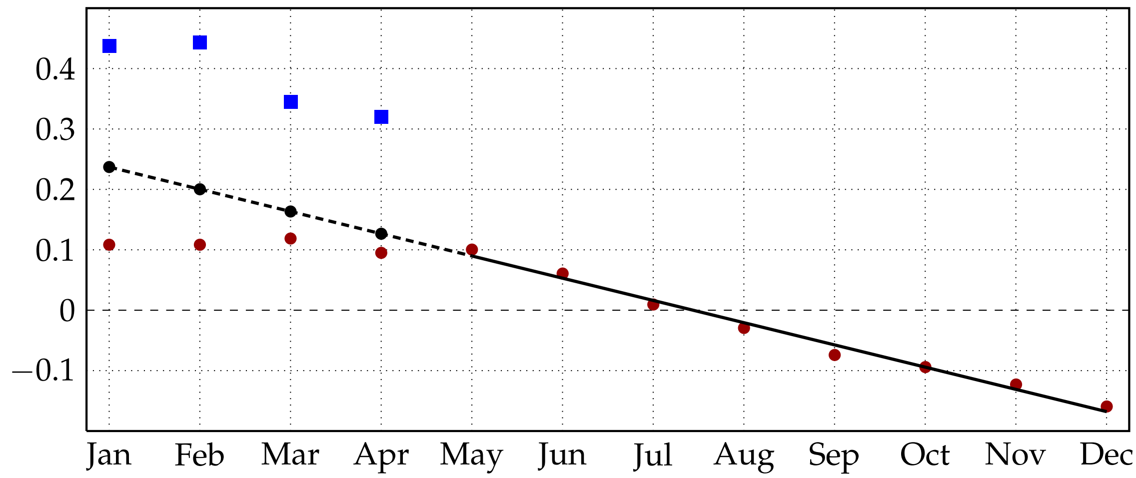
Predicting average test scores. In our second step, we compute the predicted average test scores of students born between January and April, had all students started regularly. We use the estimated coefficients $\hat{\alpha}$ and $\hat{\beta}$ from equation (1) to compute $\widehat{A}(U, R, m) = \hat{\alpha} + \hat{\beta} m$, for each $m \in \{1, \ldots, 4\}$.

Figure 2 illustrates our methodology. This figure presents again average mathematics test scores for students born in 2011. In our first step, we estimate equation (1) and obtain

[8] For notational simplicity, in what follows we drop the superscript t from $A^t(G, D, m)$.

the thick black line in Figure 2. This line fits average test scores for regular starters born between May and December. In our second step, we extrapolate this linear trend to the months between January and April. In this way, we obtain the predicted average test scores $\widehat{A}(U, R, m)$, had all students started regularly. These are shown with black circles in Figure 2.

Figure 2: Selection into early starting. Mathematics. Cohort 2011



Notes: Red circles represent average test scores for regular starters. The thick line, estimated from equation (1), fits average test scores of regular starters born between May and December. Black circles show predicted average test scores $\widehat{A}(U, R, m)$. Squares (in blue) represent the average test scores that early starters would have obtained had they started regularly, $\widehat{A}(S, R, m)$, as computed from equation (2).

Calculating counterfactual test scores for early starters. In our third step, we calculate the scores that early starters would have obtained if they had not been selected into early starting. Our methodology allows for the indirect identification of $A(S, R, m)$. The average test score $A(U, R, m)$ of all students, had all of them started regularly, is a weighted average of the scores of those selected and those not selected:

$$A(U, R, m) = P_S(m)A(S, R, m) + [1 - P_S(m)]\, A(NS, R, m).$$

where $P_S(m)$ denotes the proportion of students born in month m selected into early starting. We observe both $P_S(m)$ and $A(NS, R, m)$ in our sample. In our second step, we compute $\widehat{A}(U, R, m)$. Then, the predicted average test score $\widehat{A}(S, R, m)$ of early starters born in m can be easily expressed as

$$\widehat{A}(S,R,m) = (P_S(m))^{-1}\left[\widehat{A}(U,R,m) - (1-P_S(m))\,A(NS,R,m)\right]. \tag{2}$$

The blue squares in Figure 2 depict the predicted average test score $\widehat{A}(S,R,m)$ of early starters, as computed from equation (2).

Our measure of the strength of selection is $\widehat{A}(S,R,m) - \widehat{A}(U,R,m)$: the difference between the (predicted) average test score of early starters, had they started regularly, and the (predicted) average test score in the population, had all students started regularly. This measure is the vertical distance between the blue squares and the black circles in Figure 2. We compute the standard errors associated with this difference using bootstrap at the school level.

Our second contribution is to estimate the penalty, in terms of test scores, from starting school early. The penalty from starting early for students who are selected to start early is given by:

$$A(S,E,m) - A(S,R,m). \tag{3}$$

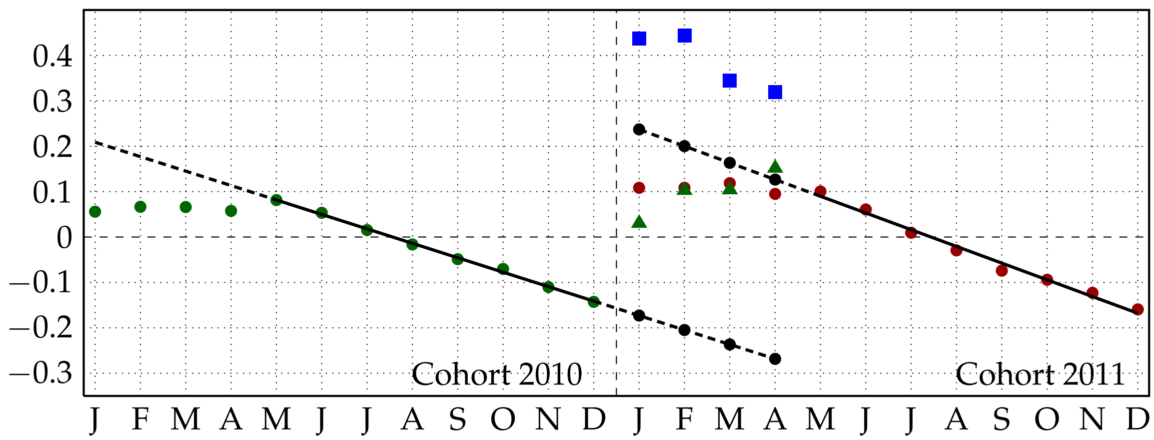
Instead, for the average student in the population, the penalty is given by:

$$A(U,E,m) - A(U,R,m). \tag{4}$$

We observe the term $A(S,E,m)$ in our data. We apply our methodology to estimate the strength of selection to obtain estimates of all other terms in equations (3) and (4).

Figure 3 illustrates how we measure the penalty from starting early. This figure presents average test scores in mathematics for two successive cohorts. The left panel shows test scores for students born in 2010. The right panel shows scores for students born in 2011. Circles represent average test scores for regular starters. Green circles on the left correspond to students born in 2010, while red circles on the right correspond to students born in 2011. Green triangles represent the actual scores $A(S,E,m)$ that early starters obtain. These students are born in 2011 and enroll in second grade in the academic year 2017–18. We compare them to the average test scores $\widehat{A}(S,R,m)$ of early starters, had they started regularly. As in Figure 2, blue squares depict $\widehat{A}(S, R, m)$. The vertical difference between the green triangles and the blue squares represents the penalty for early starters. Instead, the penalty for the average student in the population is given by the difference between the (estimated) average test score $\widehat{A}(U, E, m)$, had all students started early, and the (estimated) average test score $\widehat{A}(U, R, m)$ had all students started regularly. This difference is represented by the vertical distance between black circles in Figure 3.

Figure 3: Penalty from early starting (early starters and the average student). Mathematics. Cohort 2011



Notes: The left panel shows test scores for students born in 2010. The right panel shows test scores for students born in 2011. Green circles (for students born in 2010) and red circles (for students born in 2011) represent average test scores of regular starters. The thick line on the left panel fits average test scores of regular starters born between May and December 2010. The thick line on the right panel fits average test scores of regular starters born between May and December 2011. Squares (in blue) represent the average test scores that early starters born in 2011 would have obtained had they started regularly, as computed from equation (2). Triangles (in green) represent the actual average test scores of early starters born in 2011.

4 Results

We estimate the strength of selection following the three steps described in the previous section and illustrated in Figure 2. We first present our estimates for the effect of an extra month of age on test scores (the first step of our methodology) from equation (1). Table 1 reports these estimates for each subject and for each cohort. The estimated age-in-month effect on test scores ranges between -0.29 and -0.37.

In our second step, we directly compute the predicted average test score $\widehat{A}(U, R, m)$ in

Table 1: The age-in-month effect on test scores $\widehat{\beta}$

	2005	2006	2007	2008	2009	2010	2011
Mathematics	-0.034	-0.033	-0.034	-0.031	-0.035	-0.032	-0.037
Italian	-0.033	-0.031	-0.035	-0.031	-0.029	-0.030	-0.033

Notes: This table presents the estimates of the linear age-in-month effects β for all cohorts, for mathematics and Italian. All shown estimates are statistically significant: p-values < 0.001. Standard errors are computed using bootstrap at the school level.

the population, had all students started regularly (the black circles in Figure 2). Finally, in the third step, we use equation (2) to compute the predicted average test score $\widehat{A}(S, R, m)$ of early starters, had they started regularly (the blue squares in Figure 2). Figure 2 shows that the difference $\widehat{A}(S, R, m) - \widehat{A}(U, R, m)$ is positive for test scores in mathematics for the cohort 2011. This is evidence that early starters are positively selected for all months (January to April).

We show that there is positive selection for all cohorts and for all months. Table 2 presents the strength of selection for all months (January to April), for all cohorts (2005 to 2011), and for both subjects. Estimates are positive and significant in all cases. The estimated strength of selection ranges between 0.124 and 0.337.

We next present our results on the penalty from starting early. Figure 3 illustrates this penalty for the cohort 2011. Our estimate of the penalty for an early starter is given by the difference between $A(S, E, m)$ (the green triangles in Figure 3) and $\widehat{A}(S, R, m)$ (the blue squares in Figure 3). We represent this difference with solid black bars in Figure 4. Instead, our estimate of the penalty for the average student is given by the difference between $\widehat{A}(U, E, m)$ (the black circles over the left line) and $\widehat{A}(U, R, m)$ (the black circles over the right line in Figure 3). We represent this difference with white bars in Figure 4.

Is the penalty from starting early in general lower for students selected into early starting? Table 3 presents our estimates for the *difference* in this penalty between selected students and the average student in the population. These estimates are given by $\left[A(S, E, m) - \widehat{A}(S, R, m)\right] - \left[\widehat{A}(U, E, m) - \widehat{A}(U, R, m)\right]$. A positive difference implies that the penalty from early starting is lower for selected students. As Table 3 shows, the

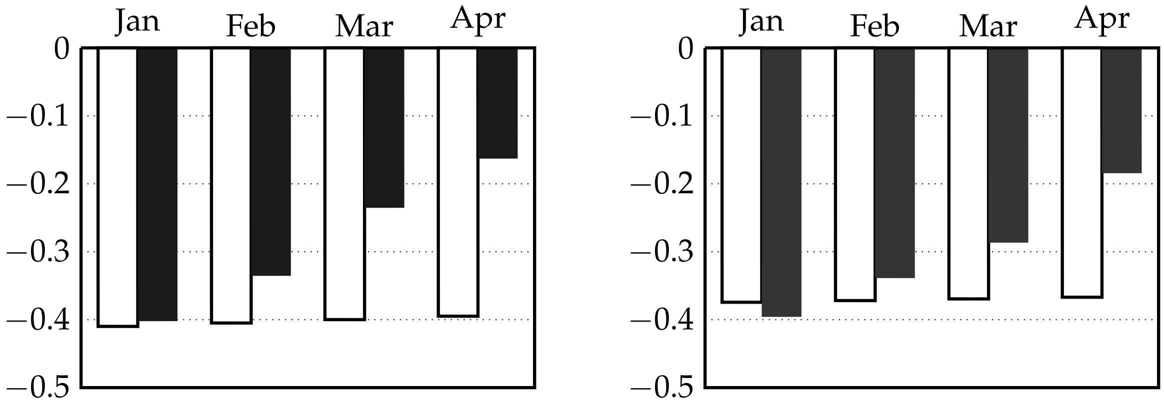
Table 2: The strength of selection

(A) Mathematics scores							
	2005	2006	2007	2008	2009	2010	2011
January	0.193	0.166	0.152	0.242	0.199	0.223	0.200
February	0.242	0.202	0.166	0.309	0.206	0.282	0.243
March	0.210	0.182	0.133	0.303	0.213	0.299	0.181
April	0.255	0.124	0.150	0.316	0.227	0.326	0.193
(B) Italian scores							
	2005	2006	2007	2008	2009	2010	2011
January	0.186	0.157	0.159	0.231	0.192	0.217	0.196
February	0.246	0.209	0.186	0.301	0.185	0.252	0.227
March	0.272	0.196	0.159	0.308	0.204	0.276	0.219
April	0.328	0.156	0.188	0.337	0.230	0.294	0.186

Notes: The strength of selection is measured by the difference between the (estimated) average test score of early starters had they started regularly and the (estimated) average test score in the population had all students started regularly. All shown estimates are statistically significant: p-values < 0.001. Standard errors are computed using bootstrap at the school level.

results are mixed for children born in January and February. There is no conclusive evidence that selected students from those months are affected by starting early differently than a random student from the population. In contrast, estimates for students born in March and April are positive and significant in 22 out of 28 cases, while non-significant in the other 6 cases. Thus, the penalty from early starting is lower for selected students born in March and April. These are the youngest students among those selected. The proportion of selected students in these months is significantly lower than in January and February.

Figure 4: Penalty from starting early (the average student vs. early starters). Cohort 2011



(a) Mathematics (b) Italian

Notes: White bars represent differences between the (estimated) average test scores in the population had all students started early (the black dots over the fitted line on the left panel in Figure 3) and the (estimated) average test scores in the population had all students started regularly (the black dots over the fitted line on the right panel in Figure 3). Solid black bars represent differences between the actual average test scores obtained by early starters born in 2011 (represented by the green triangles in Figure 3) and their (estimated) average test scores had they started earlier (blue squares in Figure 3).

5 Extensions and robustness checks

5.1 Conditioning on observable characteristics

Early starters are positively selected in all cohorts and for all months of birth. Early starters, however, have different observable characteristics than regular starters. For example, early starters are more often female and native. Moreover, their parents are more often native and more educated. Fathers of early starters are more often white-collar workers, and mothers are more often stay-at-home mothers (see Table 8 in Appendix A.4 for a comparison of all observable characteristics). Moreover, as we show next, the observable characteristics of early starters are positively correlated with better test performance. Then, is positive selection just a reflection of different observable characteristics?

We next study whether positive selection can be explained by observable characteristics. To do so, we first add observable characteristics as additional regressors to equation (1):

$$T_i^{st} = \alpha^{st} + \beta^{st} m_i^t + \gamma^{st} c_i^t + \varepsilon_i^{st} \qquad \forall\, s, t, \text{ and for } m_i \in \{5, \ldots, 12\} \tag{5}$$

Table 3: The difference in the penalty from early starting between selected and average students

	(A) Mathematics scores						
	2005	2006	2007	2008	2009	2010	2011
January	0.019*	0.001	0.037***	−0.053***	0.028***	−0.02*	0.008
	(0.073)	(0.906)	(0.000)	(0.000)	(0.008)	(0.085)	(0.466)
February	0.044***	0.061***	0.091***	−0.049***	0.085***	−0.005	0.069***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.671)	(0.000)
March	0.167***	0.166***	0.174***	0.029*	0.142***	0.008	0.165***
	(0.000)	(0.000)	(0.000)	(0.065)	(0.000)	(0.574)	(0.000)
April	0.161***	0.267***	0.221***	0.062***	0.154***	0.026	0.232***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.149)	(0.000)
	(B) Italian scores						
	2005	2006	2007	2008	2009	2010	2011
January	−0.043***	−0.02**	−0.019*	−0.057***	−0.009	−0.039***	−0.021**
	(0.000)	(0.043)	(0.070)	(0.000)	(0.378)	(0.000)	(0.024)
February	−0.058***	0.021	0.013	−0.067***	0.04***	−0.006	0.033***
	(0.000)	(0.108)	(0.253)	(0.000)	(0.003)	(0.648)	(0.009)
March	0.030**	0.100***	0.088***	−0.010	0.065***	0.024	0.083***
	(0.040)	(0.000)	(0.000)	(0.528)	(0.000)	(0.105)	(0.000)
April	−0.015	0.177***	0.121***	0.025	0.072***	0.074***	0.183***
	(0.417)	(0.000)	(0.000)	(0.139)	(0.000)	(0.000)	(0.000)

Notes: This table presents our estimates for the differences in the effect of early starting between selected students and the average student in the population. A positive difference implies that the penalty from early starting is lower for selected students. Standard errors are computed using bootstrap at the school level. P-values are reported in parentheses.

The vector c_i^t includes characteristics for student i in cohort t. These characteristics refer both to students and their parents. They include gender, whether the student or parents are foreign-born, and parents' education and labor market status. We estimate equation (5) using test scores for regular students born between May and December. We do this separately for each subject s and for each cohort t. Table 9 in Appendix A.4 reports results from estimating equation (5). We find that early starters have observable characteristics that are associated to better academic performance. See Appendix A.4 for a detailed description of the variables included in c_i^t.

Next, we compute the test scores *adjusted* by observable characteristics: $\widetilde{T}_i^{st} \equiv T_i^{st} - \widehat{\gamma}^{st} c_i^t$. The adjusted test score $\widetilde{T}_i^{st}$ measures the part of the individual score not explained by observables. We follow the methodology described in Section 3 to estimate the strength of selection, using now adjusted test scores $\widetilde{T}_i^{st}$ instead of T_i^{st}.

We find that there is positive and significant selection for all cohorts, even after controlling for observable characteristics. We report all estimates of the strength of selection controlling for observable characteristics in Table 10 in Appendix A.4.

Table 4 compares the estimates of the strength of selection with and without controls, for the 2011 cohort.[9] To compute these estimates, we use the subsample of students with information on all observable characteristics.[10] We present estimates for each month and for both subjects. Columns (I) and (III) report the estimates of the strength of selection without controls. Columns (II) and (IV) report the estimates of the strength of selection with controls. The estimates of the strength of selection with controls are only slightly lower than those without controls. Observable characteristics only explain a small fraction of the strength of selection.[11]

5.2 Regional analysis

We finally apply our methodology to identify the sign and the strength of selection at a disaggregated geographical level. Italy is divided into 20 regions aggregated into three distinct macro-regions (North, Center, and South). There is substantial heterogeneity in socioeconomic characteristics between these three macro-regions.

The map in Figure 5 shows the proportion of early starters in each Italian region for students born between January and April. The fraction of early starters is heterogeneous across areas. While in the southern region of Campania, 62.7% of students born between

[9] The results for other cohorts are qualitatively similar to those of 2011.

[10] We compute all estimates in Table 4 using this subsample. This is why the estimates of the strength of selection without controls differ from those in Table 2.

[11] We also compute the penalty from early starting using adjusted test scores. We find that the penalty for early starters is significantly lower than that for regular starters in 51 out of 56 cases. We report all estimates in Table 11 in Appendix A.4.

Table 4: The strength of selection. Estimates with and without controls. Cohort 2011

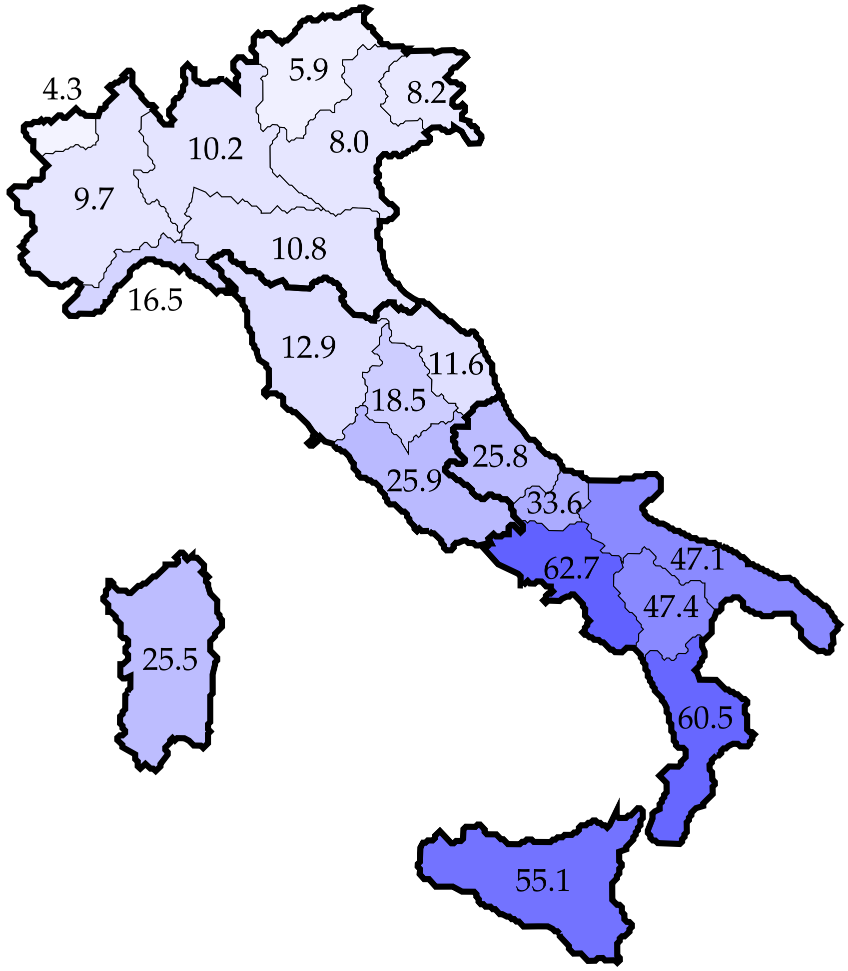
	(A) Mathematics scores		(B) Italian scores	
	(I) w/o controls	(II) w. controls	(III) w/o controls	(IV) w. controls
January	0.176	0.126	0.173	0.113
February	0.203	0.154	0.201	0.141
March	0.122	0.086	0.195	0.148
April	0.097	0.080	0.115	0.076

Notes: Columns (I) and (III) present the estimates of the strength selection, not controlling for observables. Columns (II) and (IV) present the estimates of the strength selection *controlling for observables*, that is, using adjusted test scores.

January and April start early, in the northern region of Valle d'Aosta, only 4.3% of students born in those months do so. In general, while there is also heterogeneity within each macro-region, the proportion of early starters decreases as we move from South to North. The proportion of students who start early among those born between January and April is 55.9% for the South, 22% for the Center, and 11.3% for the North.

We find that there is positive selection for all cohorts and for all months of birth in the South macro-region. In the Center macro-region, we find positive selection in January, February and March for all 42 cases except for five, where the coefficients are positive but imprecisely estimated. Most coefficients for April are imprecisely estimated. Finally, we find positive selection for students born in January in the North macro-region. Instead, most coefficients for February, March and April are imprecisely estimated. The fraction of early starters decreases not only as we move from South to North but also as we move from January to April (while 74% of students born in January start early in the South, only 3% of students born in April start early in the North). A lower proportion of early starters makes it hard to estimate the strength of selection precisely. Tables 12, 13 and 14 in the Appendix A.6 present all estimates. Hence, we conclude that the strength of selection increases as we move to the South, where the proportion of early starters is the highest.

Figure 5: Proportion of early starters by region. All cohorts



Notes: We report the fraction of early starters (for students born between January and April), for each region.

6 Discussion

How do parents decide whether their children attend preschool, which school they attend, or at which age they start formal schooling? The impact of the decisions parents make regarding their children's education may differ according to each child's characteristics. For instance, while a demanding school (or engaging in many extracurricular activities) may be beneficial for high-ability children, it may harm low-ability ones. Do parents have an accurate perception of the ability of their children?

Our paper focuses on the decision to have children start elementary school one year early. We develop an intuitive identification strategy to estimate not only the sign of selection (i.e., whether parents select high-ability children to start early) but also its strength. Our simple methodology estimates the strength of selection as the difference between the counterfactual test scores that selected students would have obtained had they *not* started early and the test scores of the average student in the population. We find robust evidence of positive selection. Early starters would have obtained test scores 0.2 standard deviations higher than the average student, had they started regularly.

Our methodology is simple and can be used to estimate the sign and strength of selection in other environments. The implementation of our methodology requires (i) a well-defined functional form that relates the outcome and running variable and (ii) an exogenous cutoff in the running variable. These conditions hold in most regression discontinuity design settings.

Our methodology also allows for the estimation of the penalty from early starting; i.e. the decrease in scores due to starting school being twelve months younger. We estimate this penalty for students who do not start early and for early starters born in different months. We find that the youngest among the earlier starters are those who suffer a (relatively) lower penalty.

We study a decision (to have children start elementary school one year early) that has several important consequences, including effects on test scores at all educational levels and labor market outcomes. Addressing whether the decision is optimal overall is beyond the scope of this paper. In this paper, we focus exclusively on test scores during compulsory education.

References

ALTON, A. AND A. MASSEY (1998): "Date of birth and achievement in GCSE and GCE A-level," *Educational Research*, 40, 105–109.

ANGRIST, J. D., E. BATTISTIN, AND D. VURI (2017): "In a small moment: Class size and moral hazard in the Italian Mezzogiorno," *American Economic Journal: Applied Economics*, 9, 216–49.

ANGRIST, J. D. AND W. N. EVANS (1998): "Children and their parents' labor supply: Evidence from exogenous variation in family size," *The American Economic Review*, 88, 450–477.

ANGRIST, J. D. AND A. B. KRUEGER (1991): "Does compulsory school attendance affect schooling and earnings?" *The Quarterly Journal of Economics*, 106, 979–1014.

BEDARD, K. AND E. DHUEY (2006): "The persistence of early childhood maturity: International evidence of long-run age effects," *The Quarterly Journal of Economics*, 121, 1437–1472.

BLACK, D. A., J. JOO, R. LALONDE, J. A. SMITH, AND E. J. TAYLOR (2022): "Simple tests for selection: Learning more from instrumental variables," *Labour Economics*, 79, 102237.

BLACK, S. E., P. J. DEVEREUX, AND K. G. SALVANES (2011): "Too young to leave the nest? The effects of school starting age," *The Review of Economics and Statistics*, 93, 455–467.

BORG, M. G. AND J. M. FALZON (1995): "Birth date and sex effects on the scholastic attainment of primary schoolchildren: A cross-sectional study," *British Educational Research Journal*, 21, 61–74.

CLARKE, D., S. OREFFICE, AND C. QUINTANA-DOMEQUE (2019): "The demand for season of birth," *Journal of Applied Econometrics*, 34, 707–723.

Cook, P. J. and S. Kang (2016): "Birthdays, schooling, and crime: Regression-discontinuity analysis of school performance, delinquency, dropout, and crime initiation," *American Economic Journal: Applied Economics*, 8, 33–57.

——— (2018): "The school-entry-age rule affects redshirting patterns and resulting disparities in achievement," NBER Working Papers 24492.

Crawford, C., L. Dearden, and E. Greaves (2014): "The drivers of month-of-birth differences in children's cognitive and non-cognitive skills," *Journal of the Royal Statistical Society Series A: Statistics in Society*, 177, 829–860.

Crawford, C., L. Dearden, and C. Meghir (2010): "When you are born matters: The impact of date of birth on educational outcomes in England," Working Paper.

Dhuey, E., D. Figlio, K. Karbownik, and J. Roth (2019): "School starting age and cognitive development," *Journal of Policy Analysis and Management*, 38, 538–578.

Dizon-Ross, R. (2019): "Parents' beliefs about their children's academic ability: Implications for educational investments," *American Economic Review*, 109, 2728–65.

Dobkin, C. and F. Ferreira (2010): "Do school entry laws affect educational attainment and labor market outcomes?" *Economics of Education Review*, 29, 40–54.

Elder, T. E. and D. H. Lubotsky (2009): "Kindergarten entrance age and children's achievement: Impacts of state policies, family background, and peers," *Journal of Human Resources*, 44, 641–683.

Fan, E., J.-T. Liu, and Y.-C. Chen (2017): "Is the quarter of birth endogenous? New evidence from Taiwan, the US, and Indonesia," *Oxford Bulletin of Economics and Statistics*, 79, 1087–1124.

Fredriksson, P. and B. Öckert (2014): "Life-cycle effects of age at school start," *The Economic Journal*, 124, 977–1004.

Gupta, A. (2018): "Nonparametric specification testing via the trinity of tests," *Journal of Econometrics*, 203, 169 – 185.

KINSLER, J. AND R. PAVAN (2021): "Local distortions in parental beliefs over child skill," *Journal of Political Economy*, 129, 81–100.

MASSEY, A., G. ELLIOTT, AND E. ROSS (1996): "Season of birth, sex and success in GCSE English, mathematics and science: some long-lasting effects from the early years?" *Research Papers in Education*, 11, 129–150.

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, 1–29.

ORDINE, P., G. ROSE, AND D. SPOSATO (2018): "Parents know them better: the effect of optional early entry on pupil's schooling attainment," *Economic Inquiry*, 56, 1678–1705.

RUSSELL, R. AND M. STARTUP (1986): "Month of birth and academic achievement," *Personality and Individual Differences*, 7, 839–846.

SHARP, C. (1995): "What's age got to do with it? A study of patterns of school entry and the impact of season of birth on school attainment," *Educational Research*, 37, 251–265.

SHARP, C. AND D. HUTCHISON (1997): *How do season of birth and length of schooling affect children's attainment at key stage 1?: A question revisited*, NFER Slough.

THOMAS, S. (1995): "Considering primary school effectiveness: an analysis of 1992 Key Stage 1 results," *The Curriculum Journal*, 6, 279–295.

WEST, J., A. MEEK, AND D. HURST (2000): "Children who enter kindergarten late or repeat kindergarten: Their characteristics and later school performance," *Education Statistics Quarterly*, 2, 21–24.

A Appendix

A.1 Institutional background

The Italian education system is divided into elementary school (grades one to five), middle school (grades six to eight) and secondary school (grades nine to thirteen). Education is compulsory between the ages of six (grade one) and sixteen (grade ten). After middle school, students follow one of three tracks of secondary schooling (technical school, lyceum, or vocational school). The first two tracks lead to a high school diploma (*diploma di maturità*). Students with this diploma can then enroll in a university or other tertiary institutions.

The school year starts in mid-September and finishes in mid-June. Enrollment in elementary school is regulated by Legislative Decree number 59, issued in February 2004. According to this law, children start elementary school the year they turn six. However, children born between January and April can start school one year in advance (the year they turn five).

A.2 Linearity

Figure 1 depicts a linear relationship between test scores and age in months for children born between May and December. We next provide further evidence for this linear relationship in our data.

The Akaike Information Criterion and the Bayesian Information Criterion study the trade-off between the goodness of fit and the complexity of a model (in terms of the number of parameters). These methods assign a score to each possible model. Then, when comparing two models, each criterion selects the one with the lowest score. The main difference between these two criteria is that the BIC imposes a larger penalty than the AIC on the number of parameters. We calculate the AIC/BIC scores for our linear model and the most flexible model with month-of-birth dummies. Both criteria select the linear functional form over the functional form with dummies for all cohorts and for both subjects, except for the AIC for the 2011 cohort.

In the main empirical exercise, we use the subsample of children born from May to December to estimate age-in-months effects on test scores. We then use the estimated age-in-months effects to compute the average test scores for children born between January and April. We perform an additional exercise that provides evidence on the precision of this prediction: we use information from June to December to predict average test scores for children born in May. We then compare our prediction against predictions from polynomials up to the fifth order (in the spirit of Gupta [2018]). We find that the linear prediction is the closest to the actual value for both subjects and for all cohorts.

A.3 Information from other grades

The INVALSI data contains information on grades two, five, eight and ten. We use data from grade two for several reasons. First, grade two is the closest to the decision of early starting. Second, the effect of age-in-months on scores is the largest for grade two; see Table 5.

Table 5: The age-in-month effect on test scores $\widehat{\beta}$. All Grades

(A) Mathematics scores							
	2005	2006	2007	2008	2009	2010	2011
Grade 2	−0.034	−0.033	−0.034	−0.031	−0.035	−0.032	−0.037
Grade 5	−0.023	−0.023	−0.022	−0.021	−0.021	−0.021	−0.021
Grade 8	−0.011	−0.013	−0.014	−0.012	−0.014	∗	∗
Grade 10	−0.006	−0.006	−0.006	−0.008	−0.007	∗	∗
(B) Italian scores							
	2005	2006	2007	2008	2009	2010	2011
Grade 2	−0.033	−0.031	−0.035	−0.031	−0.029	−0.030	−0.033
Grade 5	−0.023	−0.024	−0.022	−0.025	−0.023	−0.024	−0.025
Grade 8	−0.017	−0.017	−0.018	−0.018	−0.020	∗	∗
Grade 10	−0.009	−0.008	−0.009	−0.010	−0.010	∗	∗

Notes: This table presents the estimates of the linear age-in-month effects β for all academic years, for Italian and mathematics. All shown estimates are statistically significant: p-values < 0.001. $(*)$: INVALSI does not provide raw test scores for grades eight and ten, for cohorts 2010 and 2011.

As a robustness exercise, we show the results of our methodology for grade five instead of grade two. We do this because absent data on grade two, the best available data would be that of grade five. We also find strong evidence of positive selection. Table 6 presents the strength of selection for all months of birth, for all cohorts and for both subjects. Estimates are significant and positive in all twenty-eight cases but four (students born in April 2007 and in April 2009, for both subjects).

Our second result compares the penalty from early starting for selected students and for the average student in the population. Table 7 presents all estimates for the difference in penalty from early starting using the data from grade five. We find that the penalty from early starting is lower for selected students born in March and April. Estimates for the difference in penalty are significant and positive in all twenty-eight cases but four. The results are mixed for children born in January and February. To sum up, the results from grade five are consistent with those from grade two.

A.4 Details on conditioning on observable characteristics

In Section 5.1 we identify the sign and strength of selection after controlling for observable characteristics. We include a vector of characteristics c in the regression of test scores on month of birth. This vector contains characteristics of the students and their parents. We include students' gender and whether they are foreign-born. For each parent, we also include whether they are foreign-born, their highest degree attained, and their labor market status.

In Appendix A.1 we describe the educational categories included in the variable highest degree attained. The possible labor market statuses of parents are: unemployed, stay-at-home, white-collar (*manager, university professor, official, professional employee, freelancer, soldier, teacher, employee*), self-employed (*entrepreneur or farm owner, self-employed worker*), blue-collar (*manual worker, service sector employee or cooperative member*), and retired. We only include observations that contain information for *all* these variables. These observations represent 69.4% of the original sample ($2,279,718$ observations for mathematics and $2,267,448$ observations for Italian).

Table 8 provides descriptive statistics for the resulting sample. Column (I) describes

Table 6: The strength of selection. Grade five

	2005	2006	2007	2008	2009	2010	2011
(A) Mathematics scores							
January	0.120	0.132	0.099	0.213	0.142	0.161	0.151
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
February	0.156	0.132	0.090	0.286	0.196	0.190	0.194
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
March	0.119	0.145	0.080	0.266	0.184	0.154	0.175
	(0.000)	(0.000)	(0.011)	(0.000)	(0.000)	(0.000)	(0.000)
April	0.165	0.131	−0.024	0.232	0.050	0.212	0.175
	(0.001)	(0.005)	(0.591)	(0.001)	(0.362)	(0.000)	(0.000)
(B) Italian scores							
	2005	2006	2007	2008	2009	2010	2011
January	0.130	0.159	0.103	0.203	0.137	0.149	0.139
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
February	0.143	0.171	0.103	0.258	0.180	0.171	0.177
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
March	0.124	0.167	0.096	0.270	0.162	0.121	0.142
	(0.000)	(0.000)	(0.003)	(0.000)	(0.000)	(0.000)	(0.000)
April	0.183	0.189	0.062	0.244	0.076	0.162	0.169
	(0.000)	(0.000)	(0.208)	(0.001)	(0.128)	(0.000)	(0.000)

Notes: The strength of selection is measured by the difference between the (estimated) average test score of early starters had they started regularly and the (estimated) average test score in the population had all students started regularly. Standard errors are computed using bootstrap at the school level. P-values are reported in parentheses.

early starters. Column (II) describes regular starters born between January and April. Column (IV) describes regular starters born between May and December. Column (III) pools observations from columns (I) and (II), so it describes all children born between January and April.

Table 9 reports the coefficients $\widehat{\gamma}$ from running the regression in equation (5) for the 2011 cohort and for both subjects.[12] The reference category for highest degree attained is elementary school. The reference category for labor market status is blue-collar.

Tables 8 and 9 show that most observable characteristics of early starters correlate positively with test scores. Early starters are more often native-born, and so are their

[12]The results for other cohorts are similar, so we do not report them here.

Table 7: Difference in the penalty from early starting for selected and average students. Grade five

	(A) Mathematics scores						
	2005	2006	2007	2008	2009	2010	2011
January	0.033***	−0.042***	0.005	−0.09***	0.002	−0.032***	0.021**
	(0.002)	(0.000)	(0.642)	(0.000)	(0.839)	(0.002)	(0.039)
February	0.026**	−0.009	0.045***	−0.118***	0.06***	0.009	0.025**
	(0.033)	(0.426)	(0.000)	(0.000)	(0.000)	(0.421)	(0.037)
March	0.148***	0.066***	0.135***	−0.012	0.11***	0.08***	0.074***
	(0.000)	(0.000)	(0.000)	(0.479)	(0.000)	(0.000)	(0.000)
April	0.155***	0.134***	0.298***	0.019	0.287***	0.078***	0.109***
	(0.000)	(0.000)	(0.000)	(0.382)	(0.000)	(0.000)	(0.000)
	(B) Italian scores						
	2005	2006	2007	2008	2009	2010	2011
January	−0.01	−0.095***	0.002	−0.08***	−0.025***	−0.033***	−0.015
	(0.294)	(0.000)	(0.824)	(0.000)	(0.010)	(0.001)	(0.113)
February	−0.003	−0.086***	0.034***	−0.099***	0.027**	0.001	−0.014
	(0.831)	(0.000)	(0.008)	(0.000)	(0.039)	(0.941)	(0.278)
March	0.095***	0.034**	0.143***	−0.003	0.08***	0.073***	0.051***
	(0.000)	(0.019)	(0.000)	(0.847)	(0.000)	(0.000)	(0.001)
April	0.068***	0.036**	0.235***	0.018	0.21***	0.1***	0.074***
	(0.000)	(0.025)	(0.000)	(0.412)	(0.000)	(0.000)	(0.000)

Notes: This table presents our estimates for the differences in the effect of early starting for selected students and the average student in the population. A positive difference implies that the penalty from early starting is lower for selected students. Standard errors are computed using bootstrap at the school level. P-values are reported in parentheses.

parents. Parents are also more likely to have a university degree. Mothers of early starters are more often stay-at-home, while fathers are more often white-collar workers. Table 9 shows that all these characteristics significantly and positively correlate with test scores. Finally, early starters are more likely to be female. This is correlated to higher test scores in Italian but lower test scores in mathematics.

Early starters are positively selected even after controlling for observable characteristics. We use our estimates from equation (5) to adjust test scores by observable characteristics, as described in Section 5.1. Table 10 reports the estimates for the strength of selection for all months and cohorts, and for both subjects. The estimates of the strength of selection using adjusted test scores are positive and significant in all cases but one.

Table 8: Observable characteristics of students and their parents. All cohorts

Characteristics	January–April			May–December
	(I) Early	(II) Regular	(III) Total	(IV) Regular
Male student	0.452	0.529	0.510	0.507
Foreign-born student	0.011	0.023	0.020	0.020
Foreign-born mother	0.100	0.148	0.136	0.141
Foreign-born father	0.078	0.122	0.111	0.115
Mother. Highest degree attained				
Elementary school	0.023	0.018	0.020	0.019
Middle School	0.232	0.244	0.241	0.241
Vocational school	0.045	0.085	0.075	0.073
High school	0.407	0.418	0.415	0.417
University	0.271	0.209	0.225	0.226
Other tertiary institution	0.022	0.025	0.024	0.024
Mother. Labor market status				
Unemployed	0.060	0.056	0.057	0.059
Stay-at-home parent	0.387	0.293	0.317	0.320
White-collar	0.387	0.408	0.403	0.399
Self-employed	0.089	0.091	0.091	0.089
Blue-collar	0.076	0.151	0.132	0.132
Retired	0.001	0.001	0.001	0.001
Father. Highest degree attained				
Elementary school	0.027	0.024	0.025	0.024
Middle School	0.292	0.335	0.324	0.324
Vocational school	0.053	0.098	0.087	0.086
High school	0.395	0.376	0.381	0.385
University	0.214	0.150	0.166	0.164
Other tertiary institution	0.018	0.017	0.017	0.017
Father. Labor market status				
Unemployed	0.067	0.047	0.052	0.052
Stay-at-home parent	0.004	0.004	0.004	0.004
White-collar	0.436	0.373	0.389	0.385
Self-employed	0.241	0.249	0.247	0.245
Blue-collar	0.247	0.323	0.303	0.309
Retired	0.005	0.005	0.005	0.005
Number of observations	$367,826$	$1,085,400$	$1,453,226$	$3,093,940$

Notes: This table shows the fraction of students with each characteristic. Column (I) describes early starters. Column (II) describes regular starters born between January and April. Column (IV) describes regular starters born between May and December. Column (III) pools observations from columns (I) and (II), so it describes all children born between January and April.

Table 9: The coefficients of characteristics $\widehat{\gamma}$. Cohort 2011

	Mathematics	Italian
Male student	0.053^{***}	-0.103^{***}
Foreign-born student	-0.262^{***}	-0.239^{***}
Foreign-born mother	-0.126^{***}	-0.155^{***}
Foreign-born father	-0.182^{***}	-0.228^{***}
Mother. Highest degree attained		
Middle school	0.143^{***}	0.119^{***}
Vocational school	0.151^{***}	0.127^{***}
High school	0.313^{***}	0.283^{***}
University	0.445^{***}	0.441^{***}
Other tertiary institution	0.290^{***}	0.296^{***}
Mother. Labor market status		
Unemployed	0.053^{***}	0.060^{***}
Stay-at-home parent	0.063^{***}	0.055^{***}
White-collar	0.066^{***}	0.057^{***}
Self-employed	0.043^{***}	0.034^{***}
Retired	-0.068	0.020
Father. Highest degree attained		
Middle school	0.099^{***}	0.094^{***}
Vocational school	0.116^{***}	0.119^{***}
High school	0.251^{***}	0.245^{***}
University	0.350^{***}	0.356^{***}
Other tertiary institution	0.230^{***}	0.250^{***}
Father. Labor market status		
Unemployed	0.019^{*}	0.030^{***}
Stay-at-home parent	0.077^{**}	0.089^{**}
White-collar	0.084^{***}	0.082^{***}
Self-employed	0.075^{***}	0.048^{***}
Retired	0.058^{*}	0.036
Constant	-0.601^{***}	-0.473^{***}
Observations	$216,400$	$214,897$
R-squared	0.072	0.083

Notes: This table presents estimates for the coefficients on characteristics γ in equation (5). The coefficients are marked with * if the level of significance is between 5% and 10%, ** if the level of significance is between 1% and 5% and *** if the level of significance is less than 1%.

Moreover, we find that the strength of selection without controls is larger than that with controls. However, the magnitude of this difference is small.[13]

Finally, Table 11 reports all estimates for the difference in the penalty from early starting using adjusted test scores. The results are positive and significant in 51 out of 56 cases. This shows that the penalty from early starting is lower for selected students. This provides an even starker picture than that *without controlling for observable characteristics*. Without controlling for observables, while estimates for students born in March and April are positive and significant, the results are mixed for children born in January and February (see Table 3).

A.5 Descriptive characteristics by month of birth

We use test scores from children born between May and December to construct counterfactual test scores for those born between January and April. In practice, we extrapolate the linear trend in test scores from May to December to the months from January to April. One potential concern for our identification strategy may arise if parents choose when to have their kids. In particular, parents of different characteristics may have children in different months. Moreover, if these differences in characteristics were correlated with the decision to start early, then our methodology would not accurately measure the strength of selection.

We show that parents of children born between January and April have characteristics that are almost identical to those of parents of children born between May and December. To do so, we compare columns (III) and (IV) in Table 8. Column (III) summarizes the characteristics of students and parents for children born between January and April, while column (IV) summarizes these characteristics for children born between May and December. The difference between the proportions reported in columns (III) and (IV) is always smaller than 0.6 percentage points.[14]

[13]Table 4 shows this for the 2011 cohort. The results for other cohorts are qualitatively similar to those of 2011, so we do not report them.

[14]This subsample contains more than four and half million observations. Thus, any statistical test rejects the hypothesis that the proportions are equal, even with minimal differences.

Table 10: The strength of selection. Adjusted test scores

	(A) Mathematics scores						
	2005	2006	2007	2008	2009	2010	2011
January	0.140	0.126	0.119	0.204	0.162	0.168	0.126
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
February	0.160	0.160	0.135	0.266	0.164	0.233	0.154
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
March	0.123	0.156	0.116	0.269	0.185	0.257	0.086
	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.007)
April	0.233	0.083	0.143	0.253	0.262	0.252	0.080
	(0.000)	(0.109)	(0.000)	(0.000)	(0.000)	(0.000)	(0.054)
	(B) Italian scores						
	2005	2006	2007	2008	2009	2010	2011
January	0.134	0.105	0.119	0.177	0.133	0.171	0.113
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
February	0.170	0.172	0.149	0.251	0.122	0.225	0.141
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
March	0.193	0.158	0.148	0.295	0.156	0.269	0.148
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
April	0.276	0.130	0.191	0.283	0.211	0.263	0.076
	(0.000)	(0.012)	(0.000)	(0.000)	(0.000)	(0.000)	(0.092)

Notes: Adjusted test scores represent residuals of regressing scores on controls. The strength of selection is measured by the difference between the (estimated) residual average test score of early starters had they started regularly and the (estimated) residual average test score in the population had all students started regularly. Standard errors are computed using bootstrap at the school level. P-values are reported in parentheses.

Table 11: The difference in the penalty from early starting for selected and average students. Adjusted test scores

	2005	2006	2007	2008	2009	2010	2011
(A) Mathematics scores							
January	0.098***	0.039***	0.082***	−0.029**	0.091***	0.022**	0.111***
	(0.000)	(0.000)	(0.000)	(0.019)	(0.000)	(0.045)	(0.000)
February	0.158***	0.100***	0.135***	−0.021	0.158***	0.029**	0.191***
	(0.000)	(0.000)	(0.000)	(0.130)	(0.000)	(0.021)	(0.000)
March	0.292***	0.188***	0.206***	0.046***	0.207***	0.034**	0.297***
	(0.000)	(0.000)	(0.000)	(0.004)	(0.000)	(0.027)	(0.000)
April	0.225***	0.304***	0.246***	0.107***	0.160***	0.082***	0.386***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
(B) Italian scores							
	2005	2006	2007	2008	2009	2010	2011
January	0.021**	0.024**	0.051***	−0.019*	0.050***	0.019*	0.078***
	(0.047)	(0.015)	(0.000)	(0.097)	(0.000)	(0.083)	(0.000)
February	0.032***	0.049***	0.085***	−0.035***	0.103***	0.038***	0.138***
	(0.009)	(0.000)	(0.000)	(0.007)	(0.000)	(0.004)	(0.000)
March	0.126***	0.126***	0.139***	−0.018	0.113***	0.052***	0.176***
	(0.000)	(0.000)	(0.000)	(0.247)	(0.000)	(0.001)	(0.000)
April	0.055***	0.189***	0.163***	0.055***	0.089***	0.129***	0.316***
	(0.003)	(0.000)	(0.000)	(0.004)	(0.000)	(0.000)	(0.000)

Notes: This table presents our estimates for the differences in the effect of early starting for selected students and the average student in the population. A positive difference implies that the penalty from early starting is lower for selected students. Standard errors are computed using bootstrap at the school level. P-values are reported in parentheses.

A.6 Estimates of the strength of selection by macro-region

Tables 12, 13 and 14 present the strength of selection for all months (January to April), for all cohorts (2005 to 2011), for both subjects and for all three macro-regions.

Table 12: The strength of selection. Northern Italy

	(A) Mathematics scores						
	2005	2006	2007	2008	2009	2010	2011
January	0.226	0.130	0.171	0.144	0.200	0.213	0.172
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
February	0.178	−0.079	0.074	0.084	0.029	0.215	0.146
	(0.055)	(0.451)	(0.437)	(0.367)	(0.777)	(0.044)	(0.194)
March	−0.178	−0.449	−0.195	−0.381	−0.181	0.174	−0.371
	(0.391)	(0.016)	(0.198)	(0.010)	(0.280)	(0.331)	(0.030)
April	−0.292	−0.630	−0.210	−0.399	−0.329	0.188	−0.301
	(0.372)	(0.058)	(0.405)	(0.143)	(0.184)	(0.555)	(0.356)
	(B) Italian scores						
	2005	2006	2007	2008	2009	2010	2011
January	0.253	0.166	0.242	0.198	0.163	0.225	0.169
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.002)
February	0.306	0.065	0.290	0.254	−0.013	0.240	0.141
	(0.002)	(0.572)	(0.001)	(0.012)	(0.908)	(0.030)	(0.194)
March	0.323	−0.247	0.121	−0.069	−0.129	0.081	−0.251
	(0.106)	(0.200)	(0.445)	(0.673)	(0.464)	(0.647)	(0.156)
April	0.331	−0.198	0.233	0.073	−0.433	0.327	−0.211
	(0.306)	(0.585)	(0.364)	(0.804)	(0.114)	(0.280)	(0.499)

Notes: The sample is restricted to students residing in the North of Italy. The strength of selection is measured by the difference between the (estimated) average test score of early starters had they started regularly and the (estimated) average test score in the population had all students started regularly. Standard errors are computed using bootstrap at the school level.

Table 13: The strength of selection. Central Italy

(A) Mathematics scores							
	2005	2006	2007	2008	2009	2010	2011
January	0.169	0.146	0.180	0.182	0.213	0.254	0.274
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
February	0.248	0.053	0.297	0.257	0.193	0.388	0.292
	(0.000)	(0.458)	(0.000)	(0.000)	(0.010)	(0.000)	(0.001)
March	0.190	0.241	0.188	0.379	0.088	0.467	0.267
	(0.153)	(0.088)	(0.082)	(0.001)	(0.411)	(0.000)	(0.049)
April	0.160	0.088	0.236	0.016	0.267	0.442	0.223
	(0.386)	(0.642)	(0.146)	(0.937)	(0.125)	(0.042)	(0.343)

(B) Italian scores							
	2005	2006	2007	2008	2009	2010	2011
January	0.211	0.175	0.176	0.204	0.182	0.231	0.230
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
February	0.266	0.188	0.280	0.260	0.124	0.280	0.301
	(0.000)	(0.015)	(0.000)	(0.000)	(0.103)	(0.001)	(0.000)
March	0.292	0.294	0.185	0.405	0.023	0.380	0.249
	(0.030)	(0.037)	(0.061)	(0.001)	(0.832)	(0.001)	(0.079)
April	0.440	0.177	0.241	0.182	0.359	0.155	0.128
	(0.034)	(0.409)	(0.137)	(0.395)	(0.065)	(0.467)	(0.580)

Notes: The sample is restricted to students residing in the Center of Italy. The strength of selection is measured by the difference between the (estimated) average test score of early starters had they started regularly and the (estimated) average test score in the population had all students started regularly. Standard errors are computed using bootstrap at the school level.

Table 14: The strength of selection. Southern Italy

	2005	2006	2007	2008	2009	2010	2011
(A) Mathematics scores							
January	0.113	0.079	0.060	0.123	0.093	0.092	0.107
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
February	0.163	0.158	0.073	0.171	0.134	0.124	0.157
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
March	0.178	0.147	0.080	0.194	0.173	0.144	0.148
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
April	0.248	0.076	0.079	0.226	0.159	0.163	0.133
	(0.000)	(0.039)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
(B) Italian scores							
	2005	2006	2007	2008	2009	2010	2011
January	0.111	0.081	0.061	0.133	0.096	0.099	0.114
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
February	0.174	0.160	0.075	0.183	0.120	0.118	0.138
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
March	0.198	0.174	0.080	0.211	0.158	0.162	0.189
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
April	0.240	0.101	0.092	0.241	0.164	0.164	0.134
	(0.000)	(0.004)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Notes: The sample is restricted to students residing in the South of Italy. The strength of selection is measured by the difference between the (estimated) average test score of early starters had they started regularly and the (estimated) average test score in the population had all students started regularly. Standard errors are computed using bootstrap at the school level.