

WORKING PAPER

#639 | 2022

WWP

No country for young kids? The effects of school starting age throughout childhood and beyond

GONÇALO LIMA, LUÍS CATELA NUNES, ANA BALCÃO REIS,
MARIA DO CARMO SEABRA

No country for young kids? The effects of school starting age throughout childhood and beyond

Gonçalo Lima, Luís Catela Nunes, Ana Balcão Reis, Maria do Carmo Seabra *

February 15, 2022

Abstract

Being the youngest in a cohort entails many penalties. Using administrative data of every public-school student in Portugal, we show that although performance gains from being 1-year older fade quickly from primary education to high school, age-related penalties persist through a combination of grade retention, educational tracking and testing policies. Those that start school younger are more likely to repeat grades and ultimately drop out from school. Older entrants are more likely to enroll in scientific curricula in high school, are more successful at accessing public higher education and enroll in more selective undergraduate courses.

JEL Classification: H75, I21, J13. **Keywords:** school starting age, education, student achievement.

*Lima: goncalo.lima@eui.eu. Reis: ana.balcao.reis@novasbe.pt. Nunes: lcnunes@novasbe.pt. Seabra: carmo.seabra@novasbe.pt. All authors affiliated with Nova School of Business and Economics, at Rua da Holanda, n.1, 2775-405, Carcavelos, Portugal, except Gonçalo Lima, affiliated with the European University Institute. The authors would like to thank Diogo Pereira and Miguel Madeira Ruivo for the excellent research assistance provided in the early stage of this work. We would also like to thank participants of the 5th Lisbon Economics and Statistics of Education research workshop, the 27th Meeting of the Economics of Education Association, the 10th International Workshop on Applied Economics of Education, the 13th Annual Meeting of the Portuguese Economic Journal, the 2019 International Conference on Public Economic Theory and the 2019 Education Research Oxford Symposium whose comments greatly benefited the current version of this work. Without the anonymized administrative data provided by the Portuguese Directorate General for Education and Science Statistics (DGEEC) this work would have not been possible. We also gratefully acknowledge the support of Portugal's Foundation for Science and Technology (FCT) through grant Nova SBE/BIM/2018/.

I. Introduction

Every year millions of children around the world enter school for the first time. However, school age regulations make them initiate formal schooling at very different stages of their social, emotional and cognitive development. That these differences play a significant role in explaining individual outcomes, and academic success in particular, is a well-established empirical fact¹. Children starting school older typically reveal higher cognitive capacity, as measured by standardized achievement tests (e.g. [Bedard and Dhuey, 2006](#); [Puhani and Weber, 2007](#); [McEwan and Shapiro, 2008](#); [Elder and Lubotsky, 2009](#); [Cascio and Schanzenbach, 2016](#); [Attar and Cohen-Zada, 2018](#)).

Less clear is how given institutional features change the extent of these effects throughout a child's schooling career and onto adulthood. Measured cognitive differences tend to fade as children age, but effects may persist through other mechanisms. In countries where grade retention is a typical strategy, younger entrants are likelier to repeat grades (e.g. [McEwan and Shapiro, 2008](#)). In systems that track students into different curricular offers, older students are likelier to choose or be tracked into an academically-oriented offer ([Puhani and Weber, 2007](#); [Schneeweis and Zweimüller, 2014](#); [Attar and Cohen-Zada, 2018](#)).

Recent literature shows that age differences impact on individual and social well-being through mechanisms other than academic success. Younger students are more likely to be classified as having learning disabilities and attention deficit disorders ([Dhuey and Lipscomb, 2010](#); [Elder and Lubotsky, 2009](#); [Evans et al., 2010](#); [Mühlenweg et al., 2012](#)), are less persistent and more irritable ([Mühlenweg et al., 2012](#)), are significantly more likely to suffer from bullying or victimization ([Mühlenweg and Puhani, 2010a](#)) and less likely to hold leadership positions in high-school ([Dhuey and Lipscomb, 2008](#)). Younger entrants are also shown to have a higher propensity to commit crimes as teenagers ([Landersø et al., 2017](#)), as well as a higher likelihood of juvenile delinquency ([Cook and Kang, 2016](#)) or of being incarcerated for juvenile crime ([Dhuey et al., 2017](#)). The impact on long-term outcomes is somewhat more ambiguous. While some find a causal link between starting school later and higher wages ([Fredriksson and Öckert, 2014](#); [Kawaguchi, 2011](#)) or the likelihood of becoming a corporate CEO ([Du et al., 2012](#)), others do not find long-term effects on prime-age earnings ([Black et al., 2011](#); [Dobkin and Ferreira, 2010](#))².

We contribute to the literature by empirically estimating the impact of school starting

age on a wide range of individual outcomes, from early childhood until the end of upper secondary education. For that purpose, we use de-identified longitudinal administrative records of every student enrolled in public schools in Portugal. To the best of our knowledge, we provide the first plausibly causal evidence about the impact of school starting age on student outcomes in Portugal³, a school system with idiosyncratic characteristics: contrary to most countries, beyond a binding enrollment cutoff at 1 January, parents whose children are born as early as 16 September have leeway to legally postpone their child’s entrance in school. Our identification strategy exploits variation in school starting age around cutoff discontinuities using exact birth dates (such as in [McEwan and Shapiro, 2008](#); [Evans et al., 2010](#); [Dobkin and Ferreira, 2010](#); [Peña, 2017](#); [Attar and Cohen-Zada, 2018](#)). The use of exact birth dates enable us to avoid biases induced by seasonal patterns present in coarser measures, such as quarter or month of birth (as noted in [Buckles and Hungerman, 2013](#)). Resting on a well-identified set of falsifiable assumptions, our design is analogous to a local randomized experiment ([Lee and Lemieux, 2010](#)). Our estimates rely on local polynomial methods in accordance with the growing methodological consensus for their adequacy in regression discontinuity designs ([Gelman and Imbens, 2017](#); [Imbens and Lemieux, 2008](#))⁴. Given the longitudinal nature of our samples, and in order to allay concerns with non-compliance, we estimate local average treatment effects within the context of a fuzzy regression discontinuity design, going beyond intent-to-treat effects where possible.

We find that being 1-year older when entering school leads to significant gains in perceived cognitive capacity at grades 4, 6 and 9. In grade 4, Math and Language exam scores are, on average, higher by 0.27 and 0.36 standard deviations (σ), respectively. However, the gains on test scores from starting school later fade quickly. By the end of grade 9 we estimate the impact to be of about 0.20 of a standard deviation (σ) in Language and 0.16σ in Math. The rate at which local average treatment effects fall suggests that the findings are consistent with the hypothesis that differences in cognitive maturity when taking the test—rather than school entry postponement—are driving the results. As we only have metrics of individual outcomes measured at the same time—not at the exact same age—the estimates combine the effects of starting school age with those of age-at-test.

Despite the decline in the achievement premium, we also find that school starting age

differences persist well beyond elementary education. In a country where grade retention remains ubiquitous as a remedial strategy, older entrants have a 5 percentage points (pp) lower probability of repeating at least one grade in primary education, and 4pp by the end of grade 9. Younger students are doubly penalized, as even conditional on exam performance, younger students continue to be significantly more likely to be retained in the same grade. Furthermore, older students have a lower probability (-2pp) of dropping out from school by the end of grade 9. Significantly, intent-to-treat effects show that those eligible to start school one year later are more likely to enroll in the academic track (2pp) in high school, and conditional on being enrolled in the academic track, opt more often for scientific curricula (3pp), have higher application scores to access public higher education (0.10σ) and enroll in more selective undergraduate courses (0.13σ). On the other hand, we find no evidence of differences on demand for college seats, enrollment in STEM courses, or first-choice application success.

Section II further describes the empirical context of our findings. Section III describes the empirical strategy, Section IV details the data used in the analysis and Section V shows the empirical validity of our strategy. Finally, Sections VI, VII and VIII describe all findings in greater detail. Section IX concludes.

II. Institutional Setting

Portugal's compulsory education laws dictate enrollment in first grade to be mandatory for every child who is at least 6-years old by 15 September. Yet considerable leeway is given to parents wanting to enroll their children after this date. Students born between 16 September and 31 December of a given calendar year are deemed conditional and can still enroll if parents so require and there are available places in already created classes in school⁵. The existing rules implicitly generate a second—more binding—cutoff at 1 January, as children born in the beginning of the next calendar year enroll in the following school year. Children must thus be at least 5.7 years-old by 15 September, when starting school. Since most conditional students (86%) in Portugal are not deferred to enroll the following year, a child born in the beginning of January typically enrolls in first grade 1-year later than their peers born in the end of December.

The Portuguese school system is organized in three sequential levels: early childhood education and care, basic education and upper secondary education. Basic education covers the initial nine grades of schooling and is divided in three studying cycles, of

various lengths. First cycle comprises the four initial grades of primary education and teaching of most subjects is under the purview of a single teacher. Second cycle has a length of two academic years. The third cycle of basic education—comprising grades 7 through 9—corresponds to lower secondary education. At completion of basic education, typically at age 15, students transition to upper secondary education (high school). Upper secondary education offer is divided in a general academic and vocational pathways. In the general academic track, students can select on of four concentrations: science and technology, social and economic sciences, languages and humanities, or visual arts. On the other hand, the vocational track offers a plethora of denominations, with curricula geared toward earlier integration in the labor market. Compulsory schooling laws in the country determine that students should be enrolled in school until finishing the academic year when they turn 18 or until high school graduation if before the age of 18.

Students in the Portuguese school system are evaluated through teacher testing and national exams or assessments. Barring accommodations to specific student needs, national exams in Portuguese Language and Math are performed by every student in the system, by the end of fourth grade (until 2015) and ninth grade. Children sit through the exam at exactly the same date and time, answering the same questions. Exams are then anonymously evaluated by randomly allocated evaluator teachers, from schools other than the school in which the student is enrolled. By the end of grade 6, national assessment tests follow a similar procedure. In order to graduate from high school's general academic track, students must also sit through national exams – typically completing two course-specific exams in grade 11 and another two in grade 12, in most cases Portuguese Language and Math. Students can only gain admission to college if they have a passing grade in both grade 12 exams.

Admission to public higher education in Portugal is centrally governed. Candidates are publicly listed by the government according to candidate's ranked preferences, application scores and available capacity. Application scores combine high school GPA and exam scores. The application score depends on the college and department to which the student applies. In final application scores, high school GPA must weigh a minimum of 50 percent in the admission decision. However, each tertiary education institution can set the weight of exam scores within a band of 35 to 50 percent of the total application score.

III. Empirical Strategy

We start with the basic relationship of interest captured by the following linear model:

$$Y_{ig} = \beta_0 + \beta A_i + \mathbf{X}_i \delta + \varphi_c + \epsilon_i \quad (1)$$

Where Y_{ig} is the outcome of interest of student i measured by the end of school year g , A_i is the age of student i measured in decimal years as of 15 September in the year she first enrolled in first grade, \mathbf{X}_i is a vector of individual and family background characteristics measured during the year student i first enrolled in first grade, φ_c are cohort fixed effects and ϵ_i is an idiosyncratic error term. In this setting, β represents the marginal effect of delaying enrollment by 1-year. Nonetheless, Equation 1 does not account for school starting age being likely correlated with characteristics of students and their families that are not typically observed in the data, such as learning maturity.

To overcome endogeneity concerns, we exploit exogenous variation induced by the school starting age regulations. In Section V we present evidence that exists a sharp discontinuity at 1 January, and a kink at the 16 September cutoff. Therefore, our main identification strategy relies on comparing outcomes of students that are born before the cutoff of 1 January and those that are born on or after that same date and are induced to enroll in the following school year.

We interpret our regression discontinuity results in light of a potential outcomes framework (Hahn et al., 2001). Provided that other characteristics associated with the outcomes of interest are continuous at the cutoff and birth dates around the cutoff are as good as randomly assigned, the outcomes of those born before the cutoff provide a convincing counterfactual for those that are born after the cutoff had they enrolled one year earlier.

In order to estimate our coefficients we first construct a variable B_i with 366 unique integer values (allowing for leap years) that identify the birthday of student i in the calendar year, as in McEwan and Shapiro (2008). We standardize B_i as a distance (in days) to the cutoff of 1 January ($B_i = 0$) and make it run from 1 July ($B_i = -184$) to 30 June ($B_i = 181$), so that we have the discontinuity at about mid-range of the running variable. Based on it we define an indicator variable, $\tau_i = \mathbb{1}(B_i \geq 0)$, which identifies the values of B_i that are equal to or exceed the enrollment cutoff of 1 January.

We estimate the discontinuity using both local-linear and local-quadratic regression methods, through the following reduced form weighted least squares estimator⁶:

$$\min \sum_{i=1}^{N(h)} (Y_{ig} - \alpha_0 - \alpha\tau_i - \mathbf{f}(B_i))^2 \mathbf{K}_h(\tau_i, B_i) \quad (2)$$

Where $\mathbf{f}(B_i)$ is, depending on the specification, a piecewise linear ($\phi_1 B_i + \phi_2 \tau_i B_i$) or a piecewise quadratic ($\sum_{p=1}^2 \phi_p B_i^p + \sum_{p=1}^2 \phi_{pp} \tau_i B_i^p$) function of the running variable interacted with the cutoff. $\mathbf{K}_h(\tau_i, B_i)$ is a triangular weighting kernel function with bandwidth h , given by:

$$\mathbf{K}_h(\tau_i, B_i) = \max \left(0, 1 - \left| \frac{B_i}{h} \right| \right) \quad (3)$$

The bandwidth h here denotes the window of birth dates (B_i) to the left and to the right of the cutoff, used to estimate the coefficient of interest α , and $N(h) : \mathbb{1} \{-h \leq B_i \leq h\}$. The triangular kernel assigns zero weight to all observations outside the interval defined by the bandwidth $[-h, h]$ and positive weights to all observations within, with weights declining symmetrically and linearly as the value of the running variable gets farther away from the cutoff. In order to avoid imposing an *ad hoc* bandwidth length, we use data-driven methods to select optimal bandwidth for each regression. In particular, we use an upgraded version of the mean square error (MSE) optimal bandwidth selectors discussed in [Imbens and Kalyanaraman \(2012\)](#) (for the linear case) and [Calonico et al. \(2014b\)](#) (for the quadratic case) and derived in [Calonico et al. \(2018\)](#)⁷.

Since existing rules in Portugal provide leeway for parents to delay children's enrollment between the 16 September and 31 December (covering our control group), and since some parents may not comply with the more binding cutoff of 1 January, α captures at an intent-to-treat effect. Therefore, we estimate local average treatment effects (LATE) through a fuzzy regression discontinuity design, using the indicator of being born after the cutoff (τ_i) as an instrument for school starting age (A_i):

$$\min \sum_{i=1}^{N(h)} (A_i - \theta_0 - \theta\tau_i - \mathbf{f}(B_i) - \mathbf{X}_i\delta - \varphi_c)^2 \mathbf{K}_h(\tau_i, B_i) \quad (4)$$

$$\min \sum_{i=1}^{N(h)} \left(Y_{ig} - \beta_0 - \beta\hat{A}_i - \mathbf{f}(B_i) - \mathbf{X}_i\psi - \varphi_c \right)^2 \mathbf{K}_h(\tau_i, B_i) \quad (5)$$

Estimates of parameter θ in first stage Equation 4 identify the estimated proportion of the complier population with the delayed school entrance close to the cutoff. Estimates

of β , in Equation 5, identify the local average treatment effects for those complying with the cutoff (Imbens and Angrist, 1994), as further discussed in Section V. For precision, we also control for a vector of individual and family background characteristics (\mathbf{X}_i) and cohort fixed effects (φ_c).

Importantly, the plausibly causal interpretation of the effects rests on the assumption that parents do not systematically time births relative to the cutoff. In the next Section we provide evidence that there does not seem to be systematic manipulation of the running variable close to the cutoff. Additionally, we run balancing analysis of observable socioeconomic characteristics of the students around the cutoff. Finally, our working assumption is that precise birth timing around the cutoffs does not introduce sharp differences in unobserved characteristics that affect our outcomes of interest.

IV. Data

A. Data Description

We use a de-identified administrative dataset (MISI)⁸ containing detailed information on every student enrolled in Portugal from 2006 to 2017. A unique student identifier tracks students throughout grades, even as they change schools, allowing us to have a panel dataset of every student since they are first observed in the education system until the most recent observation⁹ We focus on non-adult students enrolled in the regular public system of education¹⁰.

We merge MISI data with a two other administrative datasets (ENEB and ENES) containing comprehensive information on student achievement school and national exams. Sitting standardized achievement tests in Portugal is not a stable policy, so we can only recover a few years of outcome data. During the period for which there is available data we can gather grade 4 national exam scores sat between 2013 and 2015, as well as grade 6 scores for the period 2012-2015 and grade 9, 11 and 12 information for the entire period of the dataset (2007-2017). Our first analytical dataset contains student-level information at grade 1 as well as student outcomes for those that sat at least one Math or Language grade 4, grade 6 or grade 9 national exam, a total of 660 573 children. For reduced form estimates of upper secondary and post-secondary outcomes, we use a second analytical dataset where over 630 thousand students are followed from grade 9 until the end of grade 12¹¹.

B. Main Variables

Outcomes (Y_{ig}). The main outcome variables include scores at Language and Math national exams by the end of grades 4, 6 and 9, as well as other course-specific and track-specific national exams by the end of grades 11 and 12. We also consider grade retention, dropout and graduation indicators. In the transition to high school we also construct dummies for if the student, conditional on having completed grade 9, decided to enroll in the general academic track and if, conditional on having opted for the general academic track, decided to pursue a science-oriented curricula in high school. Post-secondary outcomes include indicators for if the student applied to college, what was her application score, if she was rejected from going to college, enrolled in an academic degree, enrolled in a STEM course as well as the course selectivity of the college degree in which she enrolled.

Age variables. We define school starting age (A_i)—our main regressor of interest—as the exact student age as of 15 September of each school year. It measures the age at which the student is first observed in first grade. By the way it is constructed, unit variations in this variable represent a 1-year variation in the age at which the student is first enrolled in grade 1. On the other hand, the exact date of birth (B_i) is measured as a continuous variable (measured in days) representing all possible 366 days of the calendar year (see Section III). It is the running variable in our regression discontinuity strategy, from which we also extract the relevant post-cutoff indicator for the analysis (τ_i).

Student characteristics (X_i): We construct a vector of variables for several observable characteristics of students such as indicators for gender, first generation immigrant, access to a personal computer at home, reciprocity of school social support as a proxy variable for household financial constraints, unemployment status of the child’s father, and a proxy for the level of education in the household.

C. Descriptive Statistics

Table 1 documents summary statistics of student characteristics and outcomes for each full grade sample, and separately for students born 60-days before and after the 1 January cutoff¹². The groups born before and after the cutoff seem to be relatively homogeneous with respect to their observable individual characteristics, at least for a significance level of 1 percent. Prevalence of female students, students with access to computer at home, first generation immigrants, recipients of school social support or those whose father were in an unemployment situation when entering school is similar across samples, as expected

were the students randomly allocated around the cutoff. For higher levels of confidence, however, statistical differences may be found in some of the covariates. However, in Appendix Table 5 and Appendix Figure 12 we provide estimates that provide suggestive evidence of the continuity of each of these covariates at the cutoff, as is further discussed in the following section.

The Portuguese education system is characterized by having few immigrant students (2%) and relatively few households in which at least one of the parents has some sort of higher education (18-22%). A relatively high proportion of students received school social support¹³ when in grade 1 (16-39%). Differences across grade samples are explained by a steady increase in the rate of identification of students in need of social support in the most recent years of the data.

On the other hand, for all outcome variables and the main regressor, differences in means between the 60-days before and 60-days after cutoff samples are always statistically significant at a 1 percent confidence level. The size of these differences already help anticipate the size of some of our estimates. Students born 60-days after the cutoff have, on average about more 0.7 years of age when starting school, relative to those born 60-days before the cutoff. Likewise, children born after the cutoff have a 0.23 standard deviation point (σ) advantage in Math and 0.26σ in Language by the end of grade 4. Even if these differences tend to quickly dissipate from grade to grade, we still reject a null difference in unconditional achievement in both subjects by the end of grade 9. Likewise, children born just after the cutoff are less likely to repeat in all of the observed grades.

V. Empirical Validity

A. Do Parents Plan Birth Dates Strategically?

We begin by presenting evidence in support of our identification strategy. It could be the case that assignment to our virtual control and treatment groups is not as good as random if strategic parents—aware of their benefits—plan births to occur before or after the cutoff. Figure 1 presents in the top panel an histogram of the distribution of birth dates 30-days before and after the 1 January cutoff. Despite the downward trend in births around 7 to 10 days before the end of the year (Christmas period) the density of the running variable seems relatively continuous at the cutoff of 1 January, with no strong suspicion of parents timing births because of the cutoff. Due to their timing, downward and upward trends seem to be driven by Christmas holidays, possibly given expected reduction in medical

delivery and scheduled births for the period. The bottom panel of Figure 1 confirms this through a local cubic density manipulation test (Cattaneo et al., 2018; McCrary, 2008), where we can observe the shaded confidence intervals intersecting at the cutoff (p -value = 0.191, for a null hypothesis of dissimilar density 4-days before and after the cutoff)¹⁴. Given this evidence, our virtual assignment mechanism seems to hold its' validity.

B. Compliance

What is the age at which students start school in Portugal? Figure 2 depicts the average school starting age (SSA) of students within each birth date bin, measured as the distance in days relative to the 1 January cutoff ($B_i = 0$). The solid lines are fitted values from a piecewise quadratic spline. In the figure, it can be observed a small discrete jump in school starting age at the 16 September ($B_i = -107$). As discussed in Section II, there is leeway for parents to delay their child's SSA from this date onwards. SSA jumps slightly and, throughout the end of the calendar year, decreases at a relatively slower rate than before 16 September. A sharp discontinuity follows at the 1 January cutoff. Therefore, students born on or just after the latter cutoff average about 0.7 more years when starting school, relative to those born just before the cutoff. The fact that the difference in school starting age at the cutoff is less than 1-year between the two groups confirms imperfect compliance with the quasi-experimental design and motivates regression discontinuity estimates that take into account left-side non-compliance. About 14% of the students born on or after the 16 September defer entrance to the following school year. Appendix Figure 11 reveals the declining rates of compliance for our different samples. Compliant students are those that start school on the year they turn 6-years old and do not defer entrance before reaching the 1 January cutoff. On the other hand, our strategy does not hold on the 16 September cutoff as compliance with such a quasi-experiment would be extremely low to render powered estimates (i.e., the vast majority of parents does not delay their child's entrance at school).

Table 2 presents point estimates of the school starting age discontinuity at the cutoff for 30- and 60-days fixed bandwidths, as well as for a changing MSE-optimal bandwidth, for each of the grade samples (per row). The results are analogous to the first stage of our two-stage fuzzy discontinuity design (Equation 4). The first column reports estimates for a local baseline specification, where school starting age is regressed on a quadratic function of the running variable and a triangular kernel, for a bandwidth of 30-days before and

after the cutoff, separately for each grade sample. In accordance with Figure 2, we find a sharp and precisely estimated discontinuity at the cutoff. Column 2 reports results from a regression controlling for all student covariates presented in Table 1 and cohort fixed effects. Consistent with the evidence that student characteristics are continuous at the cutoff, there is no sizable change in the coefficient. For a longer 60-days bandwidth, results are more precise but tell a similar story, with no significant changes to the point estimates (Columns 3-5, Table 2). Columns 5 and 6 display results for bandwidths selected through data-driven methods. Results are the most precise, with coefficients varying between 0.69 and 0.74 across specifications and samples. Local polynomial estimates thus corroborate the raw differences in means on Table 1: Students born just after the cutoff are, on average, about 8 months older when starting school. However, among the compliant subpopulation this difference will evidently be 12 months, which allows us to identify the impact of starting school 1-year later.

C. Continuity

Is our virtual control group a reasonable counterfactual to older entrants? In Table 1 above, we present suggestive evidence that student characteristics of those born before and after 1 January are relatively identical close to the cutoff. The fact that point estimates in Table 2 do not change considerably after controlling for covariates further suggests that the groups are balanced in terms of observable characteristics.

Despite suggestive, these are yet not sufficient to convincingly show that covariates are continuous at the cutoff. As discussed in Section II, compulsory schooling rules in Portugal provide considerable leeway for parents to delay children’s entrance for students born between 16 September and 31 December. If parents that delay students at 16 September significantly differ in their characteristics from those that opt to not defer, then if not our intent-to-treat, our local average treatment effect (LATE) estimates could be biased by compositional effects.

We test if our identification strategy survives continuity concerns along predetermined characteristics of treated children by running regressions analogous to those in Equations 4 and 5. Alternatively, though, we regress each observable covariate on school starting age, having the cutoff as the excluded instrument. In Appendix Table 5, we summarize per-sample two-stage least square estimates for local-quadratic specifications, with and without student controls, and a 60-days fixed bandwidth. We find precisely estimated

null differences between complier groups, except for grade 4 sample students that receive school social support (for p-value $< .01$). However, the qualitative interpretation of these results is unequivocal: Since grade 4 students born on or after the 1 January cutoff are likelier to be recipients of school social support—a characteristic predictive of lower achievement—our estimated LATE, if positive, will be at most underestimated relative to the true average treatment effect at the cutoff.

VI. Impacts Throughout Basic Education

A. Student Achievement

ITT. Figure 3 depicts sharp discontinuities in student achievement in Math and Language within each birth date cell, as well as each grade sample. Fourth graders born right after the cutoff are expected to score substantially higher in each of the exams. These differences tend to vanish as students become older. Columns 1 and 2 of Table 3 present reduced form estimates, based on local-quadratic regressions using an MSE-optimal bandwidth specification, controlling for student characteristics and cohort fixed effects. On average, students assigned to treatment have consistently better performance. Subject-specific differences seem to be important. By the end of grade 9, while being assigned to treatment confers an estimated 0.09σ premium in Math relative to being born just before the cutoff, in Language the difference is 0.14σ . Nonetheless, left-side non-compliance is thus likely downward-biasing the causal impact estimated through our reduced form model.

LATE. Table 3 also summarizes the local average treatment effects (LATE) from regressions of SSA on Math and Language student achievement. School starting age is instrumented by the cutoff and includes a piecewise quadratic function of birth dates, our running variable. In accordance with the literature, we adjust standard errors for clustering within birth date cells. Column 3 presents per-grade estimates without any covariates. According to this baseline specification, entering school 1-year older entails an average benefit of about 28 percent of a standard deviation in grade 4 Math performance. When including student covariates and cohort fixed effects the effect size is slightly reduced to 0.27σ , as precision is improved through smaller standard errors (Column 4). Comparison between these two columns suggests that our identification strategy could allow us to forgo the inclusion of covariates without substantial changes in the estimated coefficients. Columns 5 and 6 present the effects for Language achievement. The local-quadratic specifications reveal a larger effect of about 0.36σ at grade 4, with clustered standard errors

at the birthday-level of 0.04σ . The estimated LATE decline at a relatively fast rate, especially between the end of grades 6 and 9. Whereas the estimated coefficient in Columns 4 and 6 drops to 0.26σ in Math and 0.29σ in Language by grade 6, respectively, by the end of grade 9 the local-quadratic estimate is as small as 0.16σ for Math and 0.20σ for Language.

Heterogeneity. We apply the same strategy to subsamples of our data in order to investigate suggestive evidence of heterogeneous effects. Figure 4 summarizes LATE estimates for different subsamples of a local-quadratic specification with student controls and cohort fixed effects. Top panel and bottom panel show the effect and 95% confidence intervals on grade 4, 6 and 9 Math and Language exam scores, respectively.

Gains in grade 4 Math are larger for girls by about 0.11σ relative to boys. As comparison, the gap between boys and girls in the Math exam in our grade 4 sample is of 0.13σ in favor of the former. However, for Math achievement, we fail to find statistically significant differences in the effect sizes across subgroups. Regarding Language performance, statistically significant benefits by grade 9 seem to be driven by students that are male, by those that do not benefit from social support and live in households where at least one of the parents or legal guardians as an higher education degree, suggesting that relatively well-off male students tend to benefit slightly more in Language, for entering school 1-year later. However, the precision of each of these estimates do not allow us to uncover statistically significant differences in effects across subgroups.

B. Grade Retention

ITT and LATE. Table 4 presents ITT and LATE of starting school 1-year older on the probability of repeating a grade, using local-quadratic specifications and controlling for cohort fixed effects. As with achievement, older entrants benefit substantially with respect to grade retention. Column 1 shows that students born after the cutoff have a significantly lower probability of repeating a grade at least once, in all grades. Restricting our attention to compliers (Column 4), the benefit becomes even larger. Compliant students born just after the cutoff are 5.3 percentage points (pp) less likely to have repeated at least once until grade 4, for an average of 13.7 percent among those born within 2 months before the cutoff (Table 1). ITT and LATE on grade retention display a different persistence pattern throughout time than impacts on student achievement do: Instead of continually decreasing, effects become slightly larger in magnitude by grade 6 (-7.9pp), quickly fading

by grade 9 (-4.4pp).

Given the estimated effects on student achievement, results in Columns 1 and 4 are not surprising. Indeed, decisions to retain students strongly rely on the achievement of cognitive capacity of students. As older students perform better, they are also less likely to be penalized by retention.

However, are younger children more likely to repeat in spite of lower achievement, or mostly because of it? Table 4 shows that, even controlling for Math or Language achievement (Columns 2-3, 5-6), older entrants are still significantly less likely to repeat a grade, even if the effect quickly approaches zero as students age. The results suggest that the impact of SSA on grade retention operate through other mechanisms. Younger students are thus doubly penalized, through lower achievement in Math and Language as well as a higher likelihood of retention.

Heterogeneity. Figure 5 depicts point estimates and 95% confidence intervals of the impact of SSA on the probability of grade retention in samples parsed by student observable characteristics. Reductions in repetition rates are larger for children from less educated parents (-11pp) and receivers of school social support (-8pp), whereas we cannot reject a null effect of SSA on the likelihood of repetition for children from highly educated parents in any of the grades. On the other hand, students from more disadvantaged backgrounds seem to benefit slightly more from delayed school entrance at the cutoff of 1 January in terms of their progression throughout basic education. These results are particularly relevant for the context of a country with high levels of repetition and large asymmetries in retention patterns across socioeconomic groups¹⁵.

C. Student Attainment

Differences in school starting age also have an impact on basic education attainment outcomes. Figure 6 depicts reduced form estimates and 95% confidence intervals for the estimated impact of SSA on a series of attainment outcomes, across different specifications. Because our data does not allow to observe whether high school students complied with the virtual assignment mechanism, we can only restrict our interpretation to ITT. Children eligible to enter school 1-year later are less likely (-2pp) to dropout, which compares to an overall average of 12% of students that dropout from school until grade 9.

SSA significantly impacts enrollment patterns in high school too. Older entrants are likelier to select a general academic track in high school (2pp)—a pathway selected by

72% of the students in our data. Figure 6 also shows that—conditional on having selected a general academic track in high school—students induced by the cutoff to start school 1-year later are also likelier (3.5pp) to enroll in the science and technology stream, a pathway that enables one to follow to STEM higher education courses, and selected by 59% of our sample.

VII. Impacts Throughout High School

A. *Student Achievement and Retention*

The patterns in student achievement observed throughout basic education persist until high school. Because we do not observe sufficient school years in our data, we can only estimate ITT for students that can be followed from grade 9 at most until grade 12. Although in most subjects we do not have enough statistical power to estimate precise nulls, for most subjects achievement gains tend to be statistically insignificant (see Appendix Figure 13). Nonetheless, we can safely affirm that older entrants have higher achievement in Physics and Biology by the end of grade 11, and Language and Biology of the order of between 5 to 10 percent of a standard deviation. Importantly, these results, albeit deriving from reduced forms, are lower bounds on the true LATE, as by high school the proportion of compliers in the control group that was not lost to grade retention, dropped out or moved to the vocational track is smaller than the same proportion in the treatment group.

Analogous to the impact on basic education outcomes, for students that have followed to upper secondary education, older entrants are less likely to be retained in grade. In Appendix Figure 14 we also show reduced form estimates of the impact of SSA on grade retention throughout upper secondary education. Students induced to start school 1-year later due to the 1 January cutoff will be less likely to be retained in grade.

B. *College Applications*

Do the effects of SSA persist along other margins even after high school graduation? We look into multiple outcomes on the applications of high school graduates to public colleges in Portugal. The left panel of Figure 7 depicts reduced form estimates and their respective 95% confidence intervals for a series of binary type of outcomes. SSA does not seem to impact on college seat demand. Students born just after the 1 January cutoff are not more or less likely to apply to higher education. Although with slightly less precision, we also fail

to find significant effects in rejection rates, as well as success rates in the first of the three application phases to public higher education in Portugal. Although, by the end of basic education differences in SSA seem to play a role in preferences for scientific subjects, we fail to find such differences by the end of high school. Our ITT suggest that older entrants are not more likely than their younger peers to be accepted into academic universities vis-à-vis polytechnic institutions. Likewise, no significant differences are uncovered for enrollment patterns in Science, Technology, Engineering and Mathematics (STEM) majors, between groups.

But in which other ways may older and younger college applicants be different? The right panel of Figure 7 depicts other margins through which SSA effects may persist after high school graduation. Significantly, we find that higher education candidates born just after the 1 January cutoff have higher application scores (0.1σ). Higher application scores enlarge the option set of candidates, as well as the chances of being admitted. Evidence of this same phenomenon is the fact that, because SSA effects on the high school GPA and some national exam scores in upper secondary education are still prevalent, older entrants also enroll in more selective courses. In our most conservative point estimates, students born just after the cutoff enroll in courses 0.11σ more selective than others.¹⁶

VIII. Robustness and Placebos

An important concern about our estimated LATE in Table 3 is that – despite controlling for cohort fixed effects – results may be driven by students that are retained in grade, clustered just before the cutoff. As shown before repeaters are disproportionately concentrated before the enrollment cutoff, which can bias estimates by introducing compositional effects. In order to overcome such concern, our first set of robustness checks is to restrict the main regressions to students that never repeated a grade. Appendix Table 6 presents results analogous to those in Table 3, only considering students that never repeated a grade. There are no statistically significant quantitative changes in the size of the effects, and certainly no changes in the qualitative interpretations of the restricted model. For each grade sample, even for the subset of relatively higher achieving never-repeaters, LATE estimates are identical, allowing to allay concerns with significant sample and attrition biases introduced by grade retention patterns.

A second concern is with patterns in birth dates reflecting parental characteristics that are not perceivable solely by inspecting the distribution of births across the calendar

year. As in other countries, scheduled birth-giving and hospital service adjustments cause the frequency of births to decrease during weekends in Portugal. If the enrollment cutoff coincidentally falls close to weekends then differences at the cutoff may introduce some correlation with the characteristics of parents. However, controlling for weekday indicators produces no changes in the point estimates, suggesting this is not a problem in our analysis (see Appendix Table 7).

A third source of concern relates to the method itself. Since local regression estimates are sensitive to the choice of bandwidth, the optimal data-driven bandwidth could be systematically biasing our results. The choice of bandwidth typically entails contemplating a trade-off: Opting for a larger bandwidth includes more valid observations and increases precision, however if it is too wide our local specification may not be adequate. Appendix Table 8 presents a sensitivity analysis, replicating results for our preferred specification with alternative bandwidths. We run the same analysis for 30-day bandwidth on both sides of the cutoff, larger 60-day bandwidths, as well as employing an alternative optimal bandwidth selector, allowing bandwidths to the left and right of the cutoff differ from each other (Calonico et al., 2018)¹⁷. Reassuringly, both point estimates and standard errors are stable across alternative bandwidths and identical to the ones presented in Tables 3 and 4. We can thus be confident that our estimated LATE are thus relatively robust to different bandwidth selection methods.

Finally, we examine if our reduced form results are not an artifact of our polynomial specifications. Calonico et al. (2014b) show that in the presence of misspecification, computed standard errors can uncover actually spurious effects. In order to test alternative p -values we run permutation tests in the spirit of Fisher (1935) and according to recently developed methods (Ganong and Jäger, 2018). In particular, we assume that the cutoff is drawn from a random distribution of 200 potential cutoffs. For each of the placebo cutoffs and the true cutoff we estimate ITT effects, as LATE would give us a non-negligible portion of meaningless under-powered estimates for most placebo cutoffs. We then compute a randomization-based p -value based on the distribution of these estimates. The strategy allows us to assess what is the likelihood of the true ITT effect being due to chance.

Appendix Figures 15, 16 and 17 show that—through randomization inference—our estimates are unlikely due to chance. All our estimates are at the end of the right-tail of the placebo distributions and are in line with the asymptotic p -values. We can also

perceive that the coefficients at the true cutoff (represented by the gray dash line in Figure 15), except at grade 9 Math, are a clear outlier among the placebos (spread out along each of the x -axis).

IX. Conclusion

An exogenous one-year variation in school starting age has significant effects on primary level student outcomes. Students that are induced to delay enrollment in first grade for one-year improve student performance in 4th grade national exams by about 0.3 standard deviations (σ) in Math and almost 0.38σ in Language in Portugal¹⁸. Heterogeneity across groups is also limited. Students from more disadvantaged backgrounds—i.e. that receive school social support and have less educated parents—seem to benefit slightly more in terms of achievement in Math. Older girls also benefit slightly more than older boys. In any case, delayed entrance is homogeneously beneficial to students across identified socioeconomic groups, with overlapping confidence intervals precluding us from taking further conclusions about these patterns. Importantly, we find that the cognitive premium by the end of elementary education persists across all groups, but quickly fades throughout lower and upper secondary education.

But through which causal mechanism do our local average treatment effects fade as students age? Since we do not have a second source of exogenous variation, we cannot separate the ‘age-at-test’ effects, reflecting cognitive maturity differences, from differential ‘exposure to schooling’ effects. Our results are thus best interpreted as absolute age effects. If underlying causal mechanisms in Portugal are no different than in other contexts (e.g. Crawford et al., 2007; Black and Devereux, 2011; Fredriksson and Öckert, 2014; Peña, 2017; Cornelissen and Dustmann, 2019), ‘age-at-test’ effects may tend to dominate and lead to null or even negative impacts of delayed school entrance in the long-run.

However, certain institutional mechanisms make school starting age matter to the individual through other margins besides measurable achievement differences. Students that enter school one-year later are less likely to repeat a grade, a pattern that persists well into high school. Conditional on achievement—arguably the most determinant factor in retention decisions—older entrants still are less likely to repeat. Likewise, we find that older entrants are less likely to dropout from basic education. SSA is important in yet other ways. Students predicted to be older entrants into school—even if being exposed to schooling later—are more likely to enroll in a general academic track in high school and,

conditional on it, to opt for high school concentrations dominated by scientific courses. Our intent-to-treat effects also show that older students have higher application scores to access public higher education (0.1σ) and enroll in more selective undergraduate courses (0.11σ). However, we find no evidence of differences on the demand for college seats, enrollment in STEM courses, or first-choice application success.

How relevant are these findings for policy and individual decision-making? Our empirical strategy—besides evidence that birth dates are not manipulated around the cutoff and that covariates are balanced independently of treatment assignment—gives us confidence that our LATE and ITT are at least internally valid. However, it is well known that RD estimates are local to the cutoff and that direct extrapolation requires relatively strong assumptions about the homogeneity of treatment (Imbens and Lemieux, 2008). We can affirm that, due to their reliability, our estimates fall with a high degree of confidence within a short interval of true estimates of the causal effects of delayed school entrance in regions of the running variable (in this case, birth dates) where parents can more easily choose to delay children. In this sense, parents can be relatively safeguarded that delaying school entrance of their children within reasonably close distance of the cutoff – even if not exactly at it – will, on average, yield the described benefits. Nevertheless, choice prescription needs be nuanced. Even if restricting our attention to short-run benefits, the policy response—if intended at improving social well-being – may be at odds with the optimal choice by parents. Strategic parents will tend to respond to evidence of benefits of being relatively older through delaying enrollment of their children. For the individual child this can signify an advantage in her school success that may (or may not) spillover into adult outcomes. However, variance-increasing effects of delaying entrance on social welfare may lead to relatively more unequal outcomes across children from different backgrounds. In the case of Portugal, if strategic parents become more responsive to evidence of relative gains to older children, this may lead to an advantage that can be perceived as unfair to those constrained in their choice. As the legal option to defer entrance is mostly granted to those whose children are born between the 16 September and 31 December, this leads to an unequal distribution of choice. Moreover, if even for parents of conditional children access to information and good professional judgment is unequally distributed, early enrollees—i.e. those that do not delay school entrance—may be disproportionately penalized. Taking our evidence at face-value, early enrollees will have lower achievement

and will also be more likely to repeat at least once during primary education, offsetting the potential future gain of entering one-year earlier into the labor market.

On the other hand, if school capacity constraints force parents to delay children entrance into formal schooling – another mechanism through which students may be delayed in the Portuguese education system – this too may have unintended consequences for ensuring equal opportunities. Children who start school a year later will typically remain in pre-school environments whose quality for learning will be more strongly correlated with family background. Many have been arguing that, insofar pre-schooling conditions are unequal, delaying public schooling may well reproduce and amplify such initial conditions (Deming and Dynarski, 2008). Both parents and policymakers should thus appropriately weigh costs associated with an additional year of childcare outside formal schooling environments and shorter work careers. Alternatively to changing policy in school entry laws, policymakers can also consider other *ex ante* measures, namely early childhood interventions aimed at addressing school readiness gaps across children from different socioeconomic groups.

X. References

- Altwicker-Hámori, S. and J. Köllö (2012, jul). Whose children gain from starting school later? - evidence from Hungary. *Educational Research and Evaluation* 18(5), 459–488.
- Attar, I. and D. Cohen-Zada (2018). The effect of school entrance age on educational outcomes: Evidence using multiple cutoff dates and exact date of birth. *Journal of Economic Behavior and Organization* 153(10568), 38–57.
- Bedard, K. and E. Dhuey (2006, nov). The persistence of early childhood maturity: International evidence of long-run age effects. *Quarterly Journal of Economics* 121(4), 1437–1472.
- Black, S. E. and P. J. Devereux (2011). Recent Developments in Intergenerational Mobility. In *Handbook of Labor Economics (Volume 4, Part B)*, pp. 1487–1541.
- Black, S. E., P. J. Devereux, and K. G. Salvanes (2011, may). Too Young to Leave the Nest? The Effects of School Starting Age. *Review of Economics and Statistics* 93(2), 455–467.
- Buckles, K. S. and D. M. Hungerman (2013, jul). Season of birth and later outcomes: Old questions, new answers. *Review of Economics and Statistics* 95(3), 711–724.
- Calonico, S., M. D. Cattaneo, and M. H. Farrell (2018). Optimal Bandwidth Choice for

- Robust Bias Corrected Inference in Regression Discontinuity Designs. *Unpublished*.
- Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal* 17(2), 372–404.
- Calonico, S., M. D. Cattaneo, and R. Titiunik (2014a). Robust data-driven inference in the regression-discontinuity design. *Stata Journal* 14(4), 909–946.
- Calonico, S., M. D. Cattaneo, and R. Titiunik (2014b). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica* 82(6), 2295–2326.
- Carlsson, M., G. B. Dahl, B. Öckert, and D. O. Rooth (2015). The effect of schooling on cognitive skills. *Review of Economics and Statistics* 97(3), 533–547.
- Cascio, E. U. and D. W. Schanzenbach (2016, jul). First in the Class? Age and the Education Production Function. *Education Finance and Policy* 11(3), 225–250.
- Cattaneo, M. D., A. Arbor, M. Jansson, and X. Ma (2018). Manipulation testing based on density discontinuity. *The Stata Journal* 18(1), 234–261.
- Cook, P. and S. Kang (2018). The School-Entry-Age Rule Affects Redshirting Patterns and Resulting Disparities in Achievement.
- 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(1), 33–57.
- Cornelissen, T. and C. Dustmann (2019). Early School Exposure, Test Scores, and Noncognitive Outcomes. *American Economic Journal: Economic Policy* 11(2), 35–63.
- Crawford, C., L. Dearden, and C. Meghir (2007). When You Are Born Matters: The Impact of Date of Birth on Child Cognitive Outcomes in England.
- Deming, D. and S. Dynarski (2008). The Lengthening of Childhood. *Journal of Economic Perspectives* 22(3), 71–92.
- Dhuey, E., D. Figlio, K. Karbownik, and J. Roth (2017). Age and Cognitive Development.
- Dhuey, E. and S. Lipscomb (2008, apr). What makes a leader? Relative age and high school leadership. *Economics of Education Review* 27(2), 173–183.
- Dhuey, E. and S. Lipscomb (2010, oct). Disabled or young? Relative age and special education diagnoses in schools. *Economics of Education Review* 29(5), 857–872.
- Dobkin, C. and F. Ferreira (2010, feb). Do school entry laws affect educational attainment and labor market outcomes? *Economics of Education Review* 29(1), 40–54.

- Du, Q., H. Gao, and M. D. Levi (2012, dec). The relative-age effect and career success: Evidence from corporate CEOs. *Economics Letters* 117(3), 660–662.
- Elder, T. E. and D. H. Lubotsky (2009). Kindergarten Entrance Age and Children’s Achievement: Impacts of State Policies, Family Background, and Peers. *The Journal of Human Resources* 44(3), 641–683.
- Evans, W. N., M. S. Morrill, and S. T. Parente (2010). Measuring inappropriate medical diagnosis and treatment in survey data: The case of ADHD among school-age children. *Journal of Health Economics* 29(5), 657–673.
- Fertig, M. and J. Kluge (2005). The Effect of Age at School Entry on Educational Attainment in Germany. *IZA Discussion Paper No.1507*.
- Fisher, R. (1935). *The Design of Experiments*. Oxford: Oliver and Boyd.
- Fredriksson, P. and B. Öckert (2014, sep). Life-cycle effects of age at school start. *Economic Journal* 124(579), 977–1004.
- Ganong, P. and S. Jäger (2018). A Permutation Test for the Regression Kink Design. *Journal of the American Statistical Association* 113(522), 494–504.
- Gelman, A. and G. Imbens (2017, aug). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, Published online.
- Hahn, J., P. Todd, and W. Van Der Klaauw (2001, jan). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* 69(1), 201–209.
- Herbst, M. and P. Strawiński (2016). Early effects of an early start: Evidence from lowering the school starting age in Poland. *Journal of Policy Modeling* 38(2), 256–271.
- Imbens, G. and K. Kalyanaraman (2012, jul). Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies* 79(3), 933–959.
- Imbens, G. W. and J. D. Angrist (1994, mar). Identification and Estimation of Local Average Treatment Effects. *Econometrica* 62(2), 467.
- Imbens, G. W. and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142(2), 615–635.
- Kawaguchi, D. (2011, jun). Actual age at school entry, educational outcomes, and earnings. *Journal of the Japanese and International Economies* 25(2), 64–80.
- Landersø, R., H. S. Nielsen, and M. Simonsen (2017, jun). School Starting Age and the Crime-age Profile. *Economic Journal* 127(602), 1096–1118.

- Lee, D. S. and T. Lemieux (2010, jun). Regression Discontinuity Designs in Economics. *Journal of Economic Literature* 48(2), 281–355.
- McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142(2), 698–714.
- McEwan, P. J. and J. S. Shapiro (2008). The Benefits of Delayed Primary School Enrollment: Discontinuity Estimates Using Exact Birth Dates The Benefits of Delayed Primary School Enrollment Discontinuity Estimates Using Exact Birth Dates. *The Journal of Human Resources* 43(1), 1–29.
- Mühlenweg, A., D. Blomeyer, H. Stichnoth, and M. Laucht (2012, jun). Effects of age at school entry (ASE) on the development of non-cognitive skills: Evidence from psychometric data. *Economics of Education Review* 31(3), 68–76.
- Mühlenweg, A. M. and P. A. Puhani (2010a). The Evolution of the School-Entry Age Effect in a School Tracking System. *Journal of Human Resources* 45(2), 407–438.
- Mühlenweg, A. M. and P. A. Puhani (2010b, mar). The Evolution of the School-Entry Age Effect in a School Tracking System. *Journal of Human Resources* 45(2), 407–438.
- Peña, P. A. (2017, feb). Creating winners and losers: Date of birth, relative age in school, and outcomes in childhood and adulthood. *Economics of Education Review* 56, 152–176.
- Ponzo, M. and V. Scoppa (2014). The long-lasting effects of school entry age: Evidence from Italian students. *Journal of Policy Modeling* 36(3), 578–599.
- Puhani, P. A. and A. M. Weber (2007, may). Does the early bird catch the worm? *Empirical Economics* 32(2-3), 359–386.
- Schneeweis, N. and M. Zweimüller (2014, apr). Early Tracking and the Misfortune of Being Young. *The Scandinavian Journal of Economics* 116(2), 394–428.
- Smith, J. (2009, jan). Can regression discontinuity help answer an age-old question in education? the effect of age on elementary and secondary school achievement. *B.E. Journal of Economic Analysis and Policy* 9(1).
- Suziedelyte, A. and A. Zhu (2015, apr). Does early schooling narrow outcome gaps for advantaged and disadvantaged children? *Economics of Education Review* 45, 76–88.
- Zhang, S., R. Zhong, and J. Zhang (2017). School starting age and academic achievement: Evidence from China’s junior high schools. *China Economic Review* 44(2016), 343–354.

XI. Tables

TABLE 1. DESCRIPTIVE STATISTICS - LONGITUDINAL DATASETS

Sample:	Full sample			60-days before cutoff			60-days after cutoff			Difference	
	Obs.	%	SD	Obs.	%	SD	Obs.	%	SD	Diff (p.p)	p-value
Student characteristics											
<i>Grade 4 sample</i>											
female	229,713	48.61	49.98	36,226	48.55	49.98	37,153	48.41	49.98	-0.15	0.69
First generation immigrant	229,713	2.31	15.03	36,226	2.27	14.90	37,153	2.27	14.89	-0.00	0.98
Access to computer at home	229,713	55.39	49.71	36,226	55.09	49.74	37,153	55.15	49.73	0.06	0.87
School social support (ASE)	229,713	38.76	48.72	36,226	39.19	48.82	37,153	39.71	48.93	0.51	0.15
Dad unemployed	186,954	6.93	25.40	29,393	7.15	25.77	30,076	7.05	25.59	-0.11	0.60
Household with higher education	197,286	21.67	41.20	30,934	21.09	40.79	32,048	21.57	41.13	0.48	0.14
<i>Grade 6 sample</i>											
Female	300,231	49.38	50.00	46,206	49.96	50.00	49,068	49.18	49.99	-0.78	0.02
First generation immigrant	300,231	2.49	15.58	46,206	2.34	15.11	49,068	2.57	15.84	0.24	0.02
Access to computer at home	300,231	46.09	49.85	46,206	45.64	49.81	49,068	46.40	49.87	0.76	0.02
School social support (ASE)	300,231	23.77	42.57	46,206	23.74	42.55	49,068	24.11	42.77	0.36	0.19
Dad unemployed	245,363	4.66	21.08	37,745	4.86	21.51	39,849	4.64	21.03	-0.23	0.14
Household with higher education	248,837	17.84	38.29	38,326	17.67	38.14	40,586	17.85	38.29	0.18	0.52
<i>Grade 9 sample</i>											
Female	189,057	51.51	49.98	28,760	52.21	49.95	31,189	51.37	49.98	-0.84	0.04
First generation immigrant	189,057	2.27	14.90	28,760	2.08	14.28	31,189	2.52	15.67	0.44	0.00
Access to computer at home	189,057	47.93	49.96	28,760	47.43	49.93	31,189	48.34	49.97	0.91	0.03
School social support (ASE)	189,057	16.23	36.87	28,760	16.26	36.90	31,189	16.32	36.96	0.06	0.84
Dad unemployed	157,119	3.67	18.81	23,886	3.64	18.72	25,749	3.73	18.95	0.09	0.59
Household with higher education	157,561	20.06	40.04	23,964	20.09	40.07	25,883	19.77	39.83	-0.32	0.37
	Obs.	Mean	SD	Obs.	Mean	SD	Obs.	Mean	SD	Diff.	p-value
Main variables											
School starting age (<i>A</i>)											
<i>Grade 4 sample</i>	229,713	6.25	0.31	36,226	5.97	0.39	37,153	6.62	0.14	0.65	0.00
<i>Grade 6 sample</i>	300,231	6.24	0.31	46,206	5.95	0.37	49,068	6.61	0.15	0.67	0.00
<i>Grade 9 sample</i>	189,057	6.24	0.31	28,760	5.93	0.36	31,189	6.61	0.15	0.68	0.00
Grade retention (%)											
<i>Grade 4 sample</i>	229,713	10.74	30.96	36,226	13.72	34.40	37,153	8.016	27.15	-5.70	0.00
<i>Grade 6 sample</i>	300,231	14.10	34.80	46,206	16.75	37.34	49,068	11.71	32.15	-5.04	0.00
<i>Grade 9 sample</i>	189,057	13.87	34.56	28,760	15.66	36.35	31,189	12.34	32.89	-3.33	0.00
Math performance (s.d.)											
<i>Grade 4 sample</i>	229,195	-0.05	0.99	35,906	-0.16	0.98	36,698	0.07	0.99	0.23	0.00
<i>Grade 6 sample</i>	299,206	0.04	0.96	46,067	-0.04	0.94	48,882	0.12	0.97	0.15	0.00
<i>Grade 9 sample</i>	187,990	0.05	0.96	28,616	0.02	0.96	30,980	0.07	0.98	0.06	0.00
Language performance (s.d.)											
<i>Grade 4 sample</i>	229,400	-0.04	0.99	36,179	-0.16	0.97	37,086	0.09	0.98	0.26	0.00
<i>Grade 6 sample</i>	299,712	0.04	0.96	46,119	-0.05	0.95	48,974	0.13	0.97	0.19	0.00
<i>Grade 9 sample</i>	188,929	0.04	0.98	28,737	-0.02	0.96	31,157	0.09	0.99	0.12	0.00

Notes: Full sample includes cohorts of students that entered regular education on a public school in continental Portugal between 2006 and 2013.

TABLE 2. EFFECT OF DELAYED SCHOOL ENTRY ELIGIBILITY AT 1 JANUARY ON SCHOOL STARTING AGE

Outcome: school starting age	30-days bandwidth		60-days bandwidth		MSE-optimal bandwidth	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Grade 4</i>						
τ_4	0.674 (0.038)	0.670 (0.038)	0.709 (0.026)	0.706 (0.026)	0.687 (0.013)	0.685 (0.014)
Observations	36131	36131	72821	72821	116362	109973
Bandwidth (in days)	30	30	60	60	94	89
<i>Grade 6</i>						
τ_6	0.730 (0.013)	0.730 (0.014)	0.741 (0.010)	0.739 (0.010)	0.733 (0.008)	0.732 (0.008)
Observations	47128	47128	94583	94583	105828	104147
Bandwidth (in days)	30	30	60	60	66	66
<i>Grade 9</i>						
τ_9	0.736 (0.023)	0.735 (0.023)	0.747 (0.015)	0.746 (0.016)	0.741 (0.010)	0.740 (0.010)
Observations	29504	29504	59492	59492	84979	81942
Bandwidth (in days)	30	30	60	60	84	82
Student controls	No	Yes	No	Yes	No	Yes
Cohort FEs	No	Yes	No	Yes	No	Yes

Notes: All coefficients are first stage estimates of per-grade local-quadratic regressions which include a post-cutoff indicator (τ) and a piecewise quadratic function of birth dates (B) interacted with τ . Where indicated, a triangular kernel with a 30-day, 60-day or MSE-optimal bandwidths is used. Regressions on Columns 2, 4 and 6 also control for cohort fixed effects as well as individual covariates in the form of indicator variables for gender (1 if female), immigrant status (1 if first generation immigrant), reciprocity of school social support (1 if receiver), dad's unemployment status (1 if unemployed), access to computer at home (1 if Yes), and fine-grained descriptions of the maximum level of education taken by the guardians of the child (e.g. primary education, lower secondary, bachelor degree, etc). Robust standard errors clustered at the birthday level are presented in parentheses.

TABLE 3. IMPACT OF SSA ON STUDENT ACHIVEMENT - ITT AND LATE

Outcome: student performance at	ITT		LATE			
	Math	Language	Math		Language	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Grade 4</i>						
School starting age	0.191 (0.017)	0.261 (0.017)	0.278 (0.030)	0.272 (0.024)	0.367 (0.046)	0.359 (0.040)
Observations	44660	100140	50849	49568	69393	64479
Bandwidth (in days)	37	82	42	40	57	52
<i>Grade 6</i>						
School starting age	0.187 (0.020)	0.210 (0.016)	0.243 (0.032)	0.255 (0.027)	0.285 (0.027)	0.290 (0.023)
Observations	90276	133323	83743	93364	93503	128476
Bandwidth (in days)	57	84	53	58	58	80
<i>Grade 9</i>						
School starting age	0.089 (0.021)	0.137 (0.022)	0.152 (0.042)	0.155 (0.037)	0.180 (0.032)	0.197 (0.028)
Observations	75451	65406	38384	43337	58843	83924
Bandwidth (in days)	76	66	39	43	58	84
Student Controls	Yes	Yes	No	Yes	No	Yes
Cohort Fes	Yes	Yes	No	Yes	No	Yes

Notes: All coefficients are estimates of local-quadratic regressions. Columns 1-3 present reduced form estimates. For these regressions the variable of interest is the post-cutoff indicator (τ), indicating treatment assignment. In Columns 4-6 the excluded instrument is τ . All regressions include a piecewise quadratic function of birth dates (B). A triangular kernel with an indicated data-driven MSE-optimal bandwidth is used. Where indicated, regressions also control for cohort fixed effects as well as individual covariates in the form of indicator variables for gender (1 if female), immigrant status (1 if first generation immigrant), reciprocity of school social support (1 if receiver), dad's unemployment status (1 if unemployed), access to computer at home (1 if yes), and fine-grained descriptions of the maximum level of education taken by the guardians of the child (e.g. primary education, lower secondary, bachelor degree, etc). Robust standard errors clustered at the birthday level are presented in parentheses.

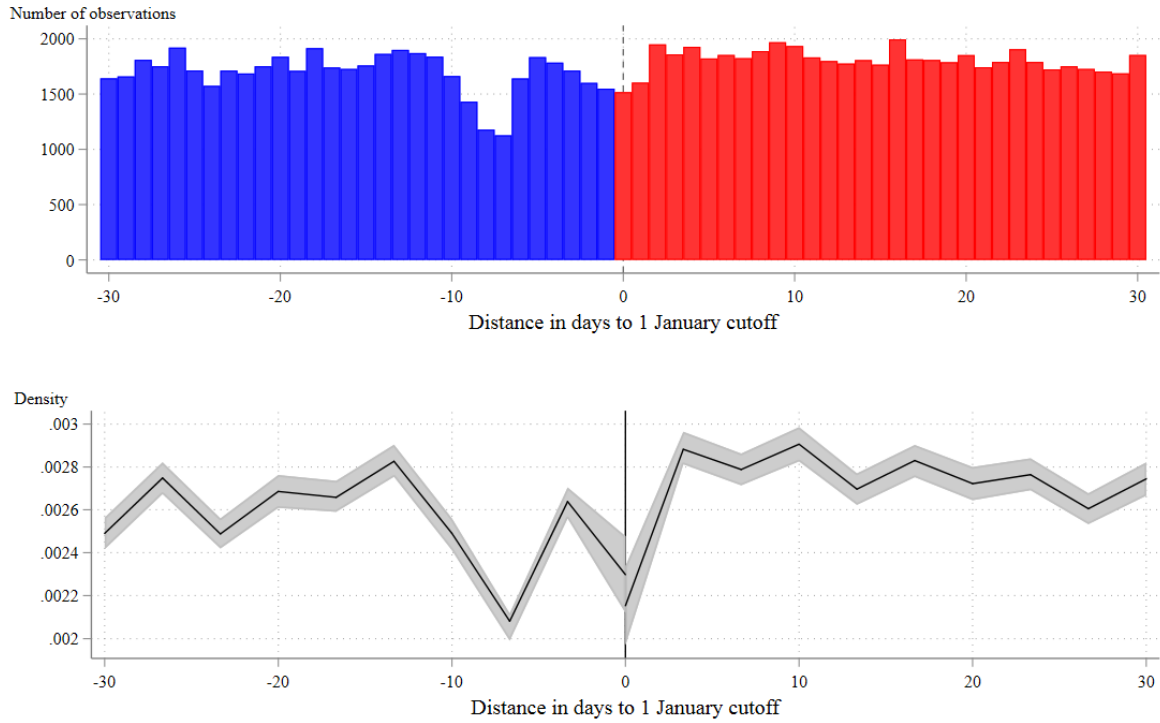
TABLE 4. IMPACT OF SCHOOL STARTING AGE ON GRADE RETENTION - ITT AND LATE

Outcome: grade retention	ITT			LATE		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Until grade 4</i>						
School starting age	-3.793 (0.751)	-2.415 (0.649)	-1.788 (0.780)	-5.275 (1.119)	-2.926 (1.094)	-2.398 (1.162)
Observations	76506	103663	78923	69498	70603	72707
Bandwidth (in days)	63	85	64	57	58	59
<i>Until grade 6</i>						
School starting age	-5.895 (0.804)	-2.932 (0.705)	-2.720 (0.720)	-7.997 (1.112)	-4.058 (1.003)	-3.695 (0.973)
Observations	90586	90276	85370	92210	80580	86997
Bandwidth (in days)	57	56	53	58	51	55
<i>Until grade 9</i>						
School starting age	-2.864 (0.778)	-1.209 (0.750)	-1.276 (0.702)	-4.397 (0.997)	-1.671 (0.985)	-2.660 (0.891)
Observations	48730	41376	49731	62472	46411	71649
Bandwidth (in days)	48	42	49	63	47	72
Math achievement control	No	Yes	No	No	Yes	No
Language achievement control	No	No	Yes	No	No	Yes
Other student controls	Yes	No	No	Yes	No	No

Notes: All coefficients are estimates of local-quadratic regressions. For these regressions the variable of interest is the post-cutoff indicator (τ), indicating treatment assignment. All regressions include a piecewise quadratic function of birth dates (B) interacted with τ , depending on the specification indicated. A triangular kernel with a data-driven MSE-optimal bandwidth is used. The excluded instrument is the post-cutoff indicator (τ). Both first and second stage regressions include a piecewise linear or quadratic function of birth dates (B) interacted with τ , depending on the specification indicated. A triangular kernel with an indicated data-driven MSE-optimal bandwidth is used. Where indicated, regressions also control for cohort fixed effects as well as individual covariates in the form of indicator variables for gender (1 if female), immigrant status (1 if first generation immigrant), reciprocity of school social support (1 if receiver), dad's unemployment status (1 if unemployed), access to computer at home (1 if yes), and fine-grained descriptions of the maximum level of education taken by the guardians of the child (e.g. primary education, lower secondary, bachelor degree, etc). Robust standard errors clustered at the birthday level are presented in parentheses.

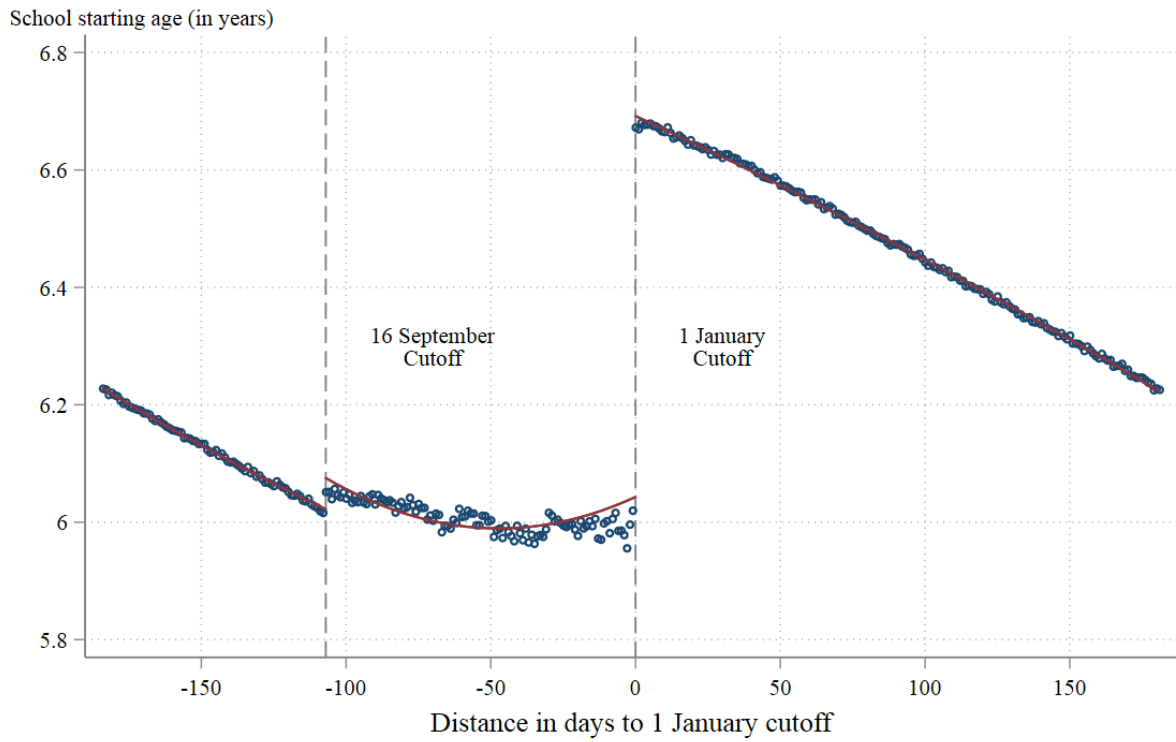
XII. Figures

FIGURE 1. DENSITY OF BIRTH DATES CLOSE TO 1 JANUARY



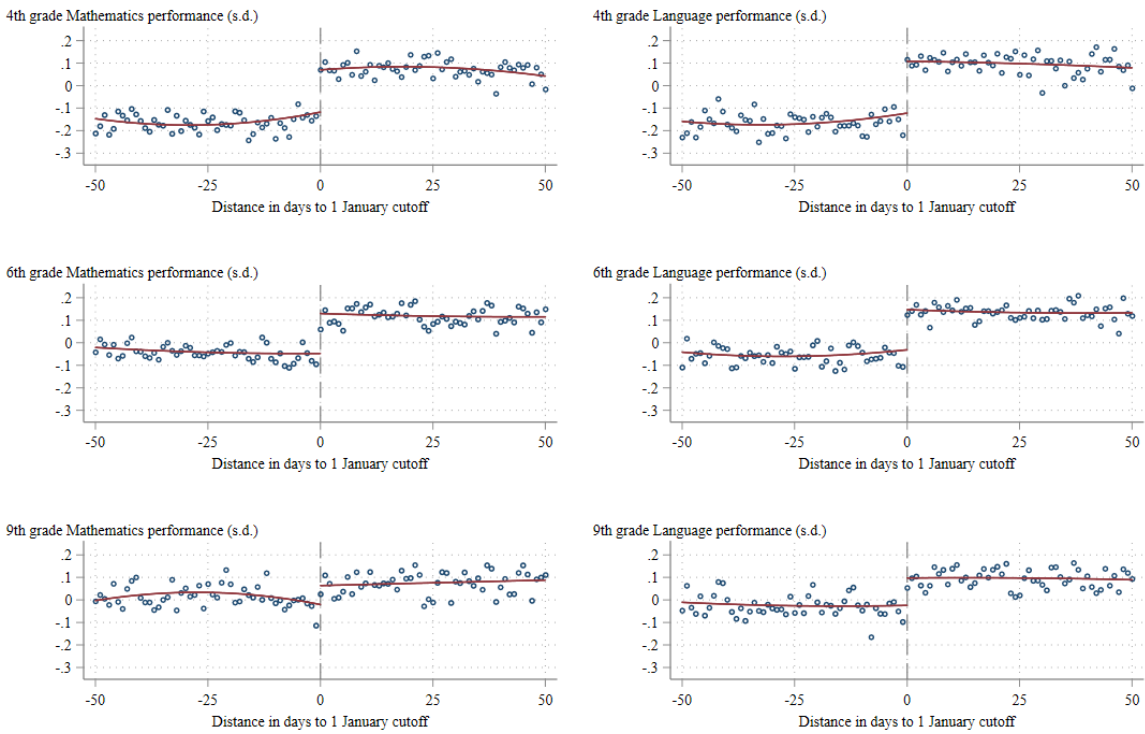
Notes: Top panel: Histogram of birth dates 30 days before and after the cutoff of 1 January. Bottom panel: Local polynomial density estimation at the cutoff, within a window of 30 days before and after 1 January. Density estimation performed through `rdensity` software package, described in Cattaneo et al. (2018). Figure is based on cohorts of students that entered a public school in continental Portugal between 2006 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in 1st grade.

FIGURE 2. DISCONTINUITIES ON SSA AT 16 SEPTEMBER AND 1 JANUARY



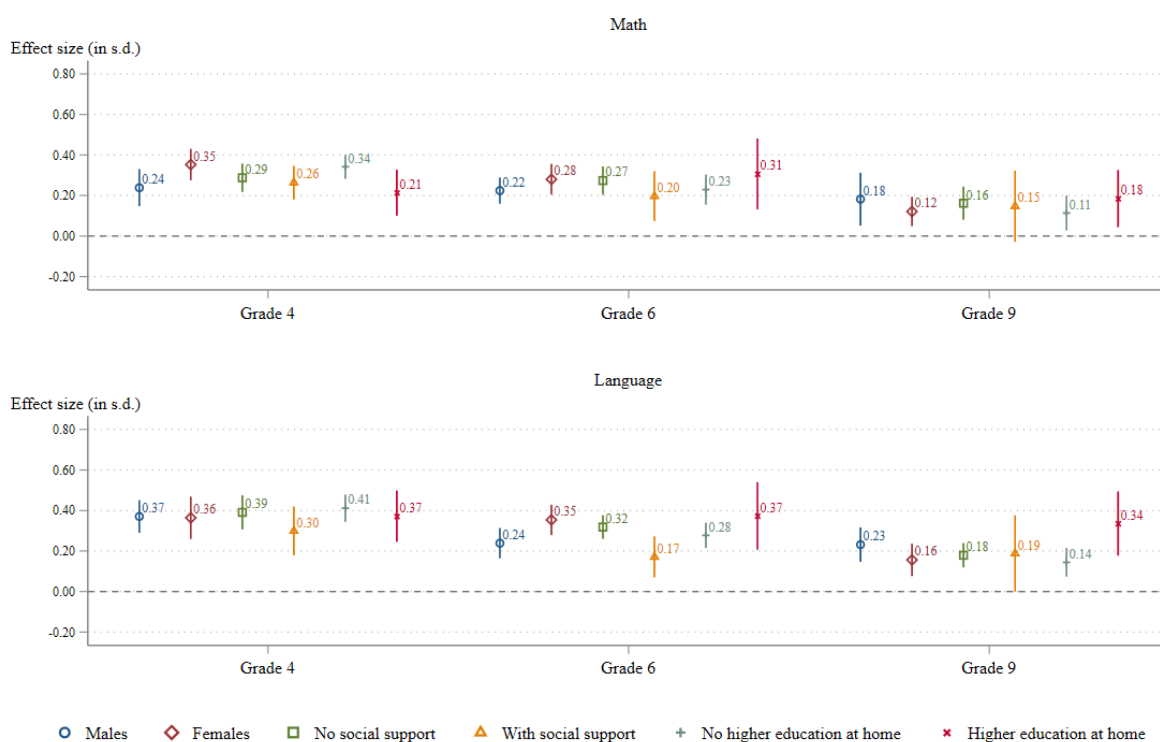
Notes: Figure is based on cohorts of students that entered a public school in continental Portugal between 2006 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in 1st grade. Horizontal axis represents the birth date relative to the cutoff date of 1 January. Each hollow circle represents within birth day cell averages of school starting age. Solid lines represent fitted values from a piecewise quadratic spline. Vertical dashed lines identify 16 September ($B_i = -107$) and 1 January ($B_i = 0$).

FIGURE 3. DISCONTINUITIES ON MATH AND LANGUAGE STUDENT ACHIEVEMENT AT 1 JANUARY, PER GRADE



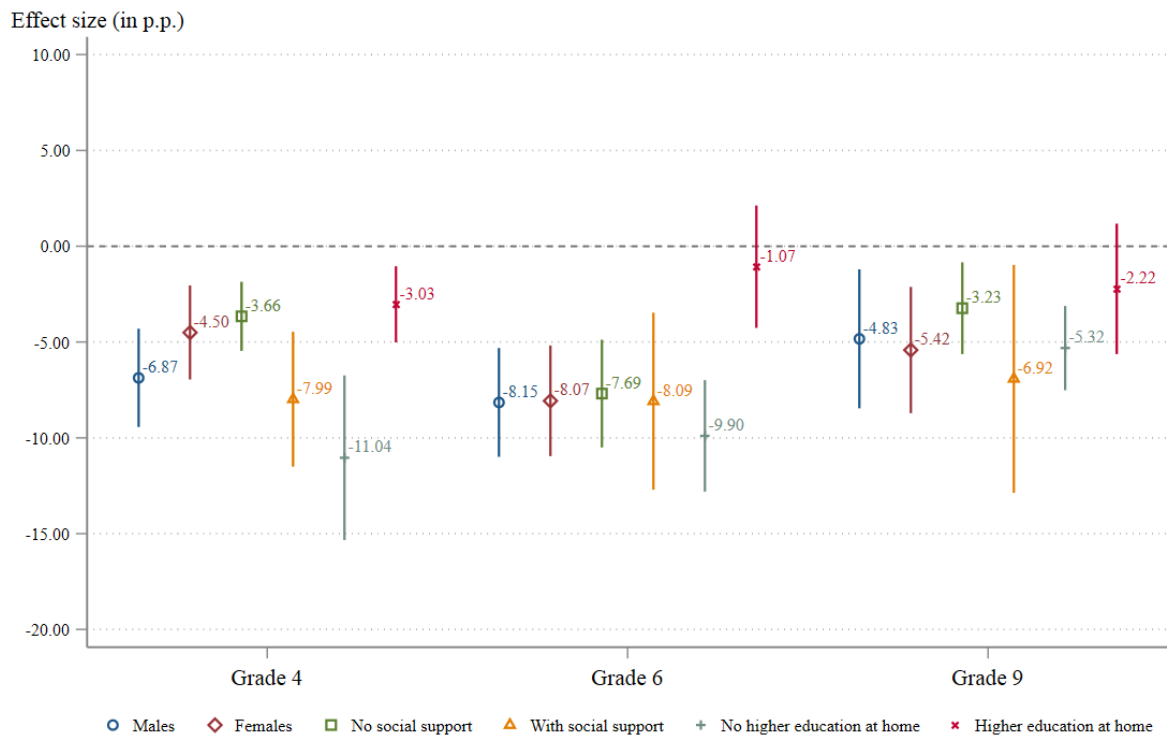
Notes: Figure is based on cohorts of students that entered a public school in continental Portugal between 2006 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in grade 1. Horizontal axis represents the birth date relative to the cutoff of 1 January. Each hollow circle represents within birth day cell averages of Math performance (left panels) and Language performance (right panels) in grade 4 (top panels), grade 6 (middle panels) and grade 9 (bottom panels). Solid lines represent fitted values from a piecewise quadratic spline. Vertical dashed lines identify 1 January ($B_i = 0$).

FIGURE 4. HETEROGENEOUS IMPACTS OF SSA ON STUDENT ACHIEVEMENT



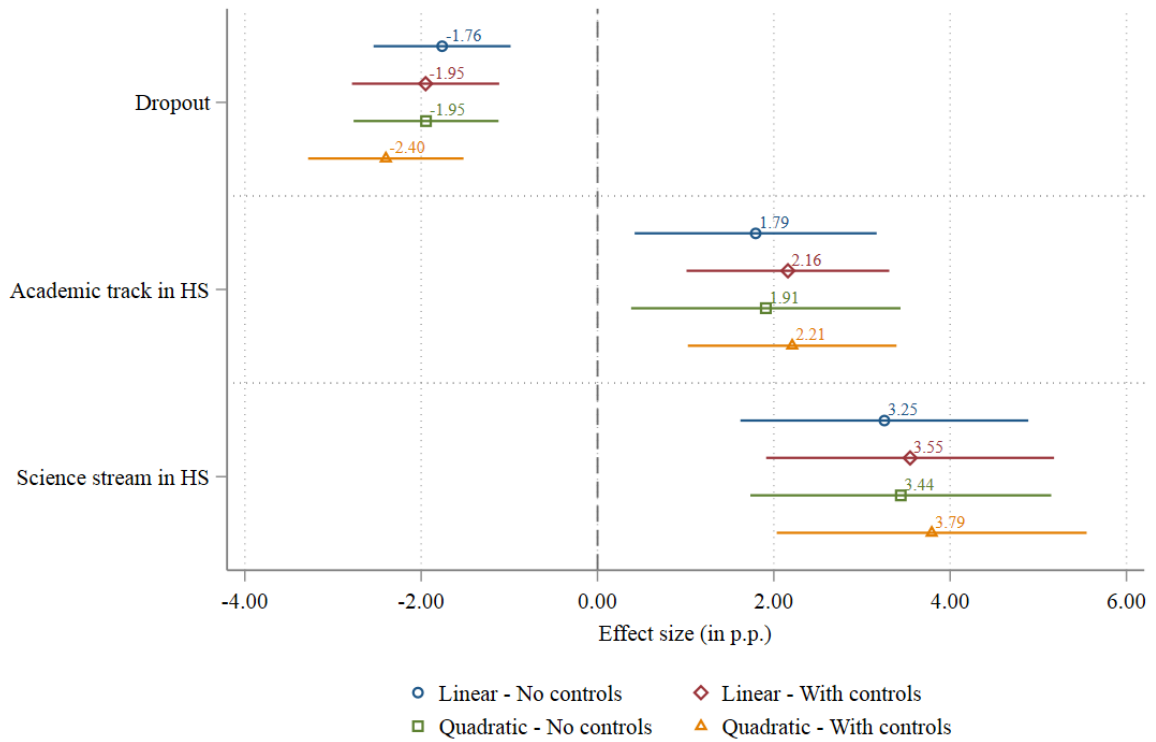
Notes: Figure is based on cohorts of students that entered a regular curriculum program in a public school in continental Portugal between 2006 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in grade 1. Each marker represents a point estimate of the impact of school starting age on student performance in grades 4, 6 and 9 Math exams (top panel) and grades 4, 6 and 9 Language exams (bottom panel) for each indicated subsample of students. Point estimates are coefficients of local regressions. The excluded instrument is the post-cutoff indicator (τ). Both first and second stage regressions include a piecewise quadratic function of birth dates (B) interacted with τ , depending on the specification indicated. A triangular kernel with data-driven optimal bandwidths is used. Regressions also control for cohort fixed effects as well as all other individual covariates. Vertical bars represent 95% confidence intervals for clustered standard errors at the birthday-level.

FIGURE 5. HETEROGENEOUS IMPACTS OF SSA ON THE LIKELIHOOD OF GRADE RETENTION



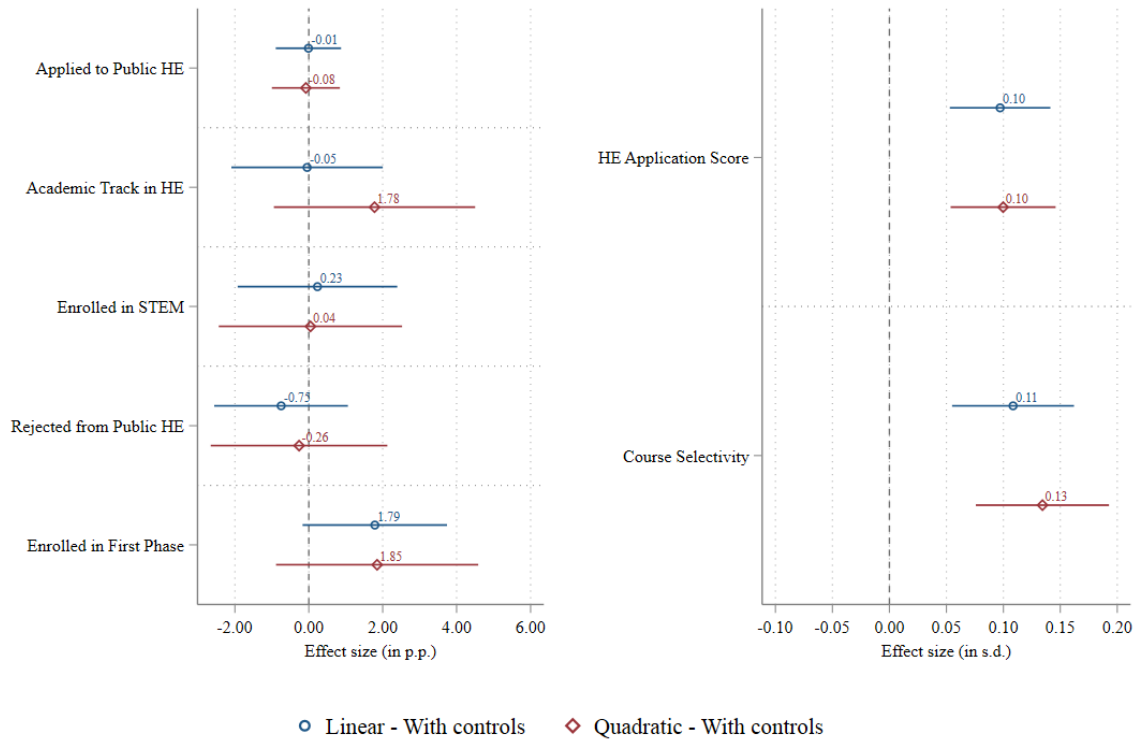
Notes: Figure is based on cohorts of students that entered a regular curriculum program in a public school in continental Portugal between 2006 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in grade 1. Each marker represents a point estimate of the impact of school starting age on grade retention until grades 4, 6 and 9. Point estimates are coefficients of local regressions. The excluded instrument is the post-cutoff indicator (τ). Both first and second stage regressions include a piecewise quadratic function of birth dates (B) interacted with τ , depending on the specification indicated. A triangular kernel with data-driven optimal bandwidths is used. Regressions also control for cohort fixed effects as well as all other individual covariates reported in Table 1. Vertical bars represent 95% confidence intervals for clustered standard errors at the birthday-level.

FIGURE 6. IMPACT OF SSA ON STUDENT ATTAINMENT



Notes: Figure is based on cohorts of students that entered a regular curriculum program in a public school in continental Portugal and were at least 5 years-old and at most 8 years-old when first enrolled in grade 1. Each marker represents a point estimate of the impact of school starting age on the probabilities of dropping out from school, opting for an academic track in high school and, conditional on having opted for an academic track in high school, having selected the science and technology stream. Point estimates are coefficients of local regressions, where the post-cutoff indicator (τ) is the regressor of interest. Regressions include a piecewise linear or a piecewise quadratic function of birth dates (B) interacted with τ , depending on the specification indicated. A triangular kernel with data-driven optimal bandwidths is used. Regressions also control for cohort fixed effects as well as all other individual covariates reported in Table 1. Horizontal bars represent 95% confidence intervals for clustered standard errors at the birthday-level.

FIGURE 7. IMPACT OF SSA ON COLLEGE APPLICATION OUTCOMES



Notes: Figure is based on cohorts of students that entered a regular curriculum program in a public school in continental Portugal and were at least 5 years-old and at most 8 years-old when first enrolled in grade 1. Point estimates are coefficients of local polynomial regressions. The regressor of interest is the post-cutoff indicator (τ). In the left panel the dependent variables, where indicated, are dummies switched on for if the students applied to a public college, enrolled in the academic track in college, enrolled in a STEM course, failed to enroll in any course due to rejection and was enrolled in application phase 1. In the right panel the dependent variables, where indicated, are standardized variables for college application scores and an index of course selectivity. Course selectivity is measured as an index of *i*) percentile rank of the pair college-course in terms of the application scores of accepted candidates, *ii*) percentile rank of the standard deviation of application scores of accepted candidates, and *iii*) acceptance rate of applications of each course in each higher education institution. The values of the latent variable are predicted through principal factor analysis, with results being later standardized to have mean zero and standard deviation of one. All regressions include a piecewise quadratic function of birth dates (B) interacted with τ . A triangular kernel with data-driven optimal bandwidths is used. Regressions also control for cohort fixed effects as well as all other individual covariates in Table 1. The horizontal bars represent 95% confidence intervals for clustered standard errors at the birthday-level.

Notes

¹The literature on the effects of school starting age emerges from such diverse contexts as the United States (e.g. Dhuey and Lipscomb, 2010; Elder and Lubotsky, 2009; Evans et al., 2010; Dobkin and Ferreira, 2010; Cascio and Schanzenbach, 2016; Cook and Kang, 2018), Sweden (Fredriksson and Öckert, 2014; Carlsson et al., 2015), Norway (Black et al., 2011), Denmark (Landersø et al., 2017), Germany (Fertig and Kluge, 2005; Puhani and Weber, 2007; Mühlenweg and Puhani, 2010a,b; Mühlenweg et al., 2012), Austria (Schneeweis and Zweimüller, 2014), the UK (Crawford et al., 2007), Poland (Herbst and Strawiński, 2016), Hungary (Altwickler-Hámori and Köllö, 2012), Chile (McEwan and Shapiro, 2008), Canada (Smith, 2009), China (Zhang et al., 2017), Japan (Kawaguchi, 2011), Australia (Suziedelyte and Zhu, 2015), Israel (Attar and Cohen-Zada, 2018), Italy (Ponzo and Scoppa, 2014) or Mexico (Peña, 2017). Bedard and Dhuey (2006) present cross-country evidence.

²For Norway, Black et al. (2011) find that the effect of school starting age becomes insignificant from 30-years of age onwards, culminating in a negative discount present value of lifelong earnings gain from starting school 1-year later. In Sweden, Fredriksson and Öckert (2014) find that being 1-year older increases educational attainment and changes labor supply over the life-cycle, with initial losses due to later entry into the labor market being offset by increased earnings after the age of 55. However, contrary to the average, individuals with low-educated parents and women tend to have higher prime-age earnings in response to school starting age. All-in-all, discounted life-time earnings are shown to fall by almost 1 percent in response to being 1-year older when enrolling in first grade.

³The exception being the estimates in Bedard and Dhuey (2006), in a cross-country context.

⁴Most recently, for instance, Cook and Kang (2016) use a local-linear estimator to uncover the causal effect of school starting age on crime outcomes. However, most studies on the impact of school starting age on academic outcomes have not relied on these type of methods. Most have focused on exploring global – rather than local – polynomial specifications (e.g. McEwan and Shapiro, 2008; Black et al., 2011).

⁵Enrollment for conditionals is not the default option; an enrollment requirement by parents is a necessary condition to starting the school year before turning 6-years old. However, it is not a sufficient one, as it also depends on school capacity constraints.

⁶Gelman and Imbens (2017) show why regression discontinuity estimation through local low-order polynomial approximations should be preferred to global polynomial regressions.

⁷Estimation, including optimal bandwidth selection, is implemented through the software package `rdrobust` developed in Calonico et al. (2014a) and Calonico et al. (2017).

⁸MISI data is collected and maintained by the Directorate General of Education and Science Statistics (*Direcção-Geral de Estatísticas da Educação e Ciência* - DGEEC), a department under the indirect administration of the Ministry of Education in Portugal.

⁹A student's track is lost when she moves abroad, drops from the education system altogether, or dies. We may also lose track of students if these move to a different public or private school and the matching algorithm is unable to correctly assign the unique identifier to new instances of the same student in the system.

¹⁰We exclude from the analysis all students enrolled in second-chance programmes, adult, vocational

education and training or artistic courses outside the scope of the regular curriculum between grades 1 and 12.

¹¹Appendix A documents in more detail all data treatment, including attrition rates for each of the cohorts.

¹²The choice of a 60-days window to each side of the cutoff is justified by full coverage of the optimal bandwidths used for estimation later on.

¹³School social support (ASE) in Portugal is tied to household financial constraints. Students identified as ASE have half- to fully-reduced price meals at school, textbooks and school material.

¹⁴The relatively sharp decline in density before the cutoff occurs during the Christmas period, with parents appearing to opt-out from having children during the period. Despite making a full analysis of the parental preferences for birth dates being out of our scope, an analysis of the histogram of births across the whole year reveals clear seasonality in the data. In particular, we find that parents prefer to have children during September, 20 September being the most common birthday in our data.

¹⁵For instance, in our analytical grade 4 sample about 13% of the children born to parents without higher education are retained in grade at least once until grade 4, while only about 1% of the children of highly educated households ever repeat a grade.

¹⁶Our index of course selectiveness takes into account courses in each higher education institution. It is constructed from the percentile rank of the mean of the application scores of accepted applicants, the percentile rank of the standard deviation of application scores of accepted candidates and acceptance rate of applications of each course in each higher education institution. The Appendix details the computation of this index.

¹⁷Optimal bandwidth selectors are deployed through the `rdrobust` software package, developed in Calonico et al. (2014a) and Calonico et al. (2017).

¹⁸The local average treatment effect of school starting age on test scores here presented are consistent with comparable evidence from other countries. Cross-country evidence from Bedard and Dhuey (2006) find effect sizes ranging from 0.2σ to 0.5σ . McEwan and Shapiro (2008)—with an identical empirical strategy, but different estimation procedure—find increases of 4th grade test scores of 0.3 - 0.4σ , in Chile. For fifth grade students in Israel, Attar and Cohen-Zada (2018) estimate somewhat smaller 0.29σ gains for Language, and 0.16σ for Math. For fourth graders in Germany, Puhani and Weber (2007) find an increase of 0.4σ .

A. Appendix

A. *The data*

For our analytical datasets, all information in period 2013-2015 for both Grades 4 and Grade 6 students is kept, whereas achievement information for Grade 9 comes from exams sat during 2016 and 2017. For Grade 4 students sitting exams in 2013-2015 we manage to follow seven different cohorts, starting school between 2006 and 2012. For each of these cohorts we can follow over 95 percent of the children between Grade 1 and Grade 4. We follow five cohorts of Grade 6 sample students entering school in the period 2006-2010, with attrition rates lower than 10 percent. Finally, given our lower bound on the data, we can follow more than 90 percent of two cohorts students until they sit Grade 9 exams in either 2016 or 2017. As we track students from their first school year to later periods in our dataset we lose more students either to private schools, a foreign country, registration failures, other non-regular education programs, or due to idiosyncratic errors in the matching algorithm.

The first analytical dataset has a total of 660 494 individual students, born between 1998 and 2008, having started school between calendar years 2006 and 2013. The distribution of information on their outcomes is as follows. Grade 4 sample: 229 661 (35%). Grade 6 sample: 300 182 (45%). Grade 9 sample: 188 648 (29%). 10% of the students have scores for both Grades 4 and 6. 28% have scores for both Grades 6 and 9. None has observed scores for all grades.

A second analytical dataset is constructed for reduced form estimates of upper secondary outcomes and out-of-sample estimation of basic education outcomes. Since, in this case, we do not require to observe each student at Grade 1, we estimate intent-to-treat effects from a total of 2 227 679 individuals born between 1983 and 2010. For an analysis of upper secondary and post-secondary outcomes, we further restrict this dataset to one where each student is followed from Grade 9 until the end of Grade 12. The restricted sample follows a total of 634 827 individuals born between 1988 and 2000, from which 171 319 (27%) applied to public higher education after graduating from high school. In the case of college application outcomes, we benefited from an hand-collected dataset, that was available on-line for a given period of time and allowed to link the data with the existing administrative datasets. We also trim the dataset for students with outlier ages by only keeping children that are at least 5-years old and had at most 8-years old when

first enrolled in Grade 1.

Course selectivity is measured as an index of *i*) percentile rank of the pair higher education institution course in terms of the application scores of accepted candidates, *ii*) percentile rank of the standard deviation of application scores of accepted candidates, and *iii*) acceptance rate of applications. The values of the latent variable are predicted through principal factor analysis, with results being later standardized to have mean zero and standard deviation of one.

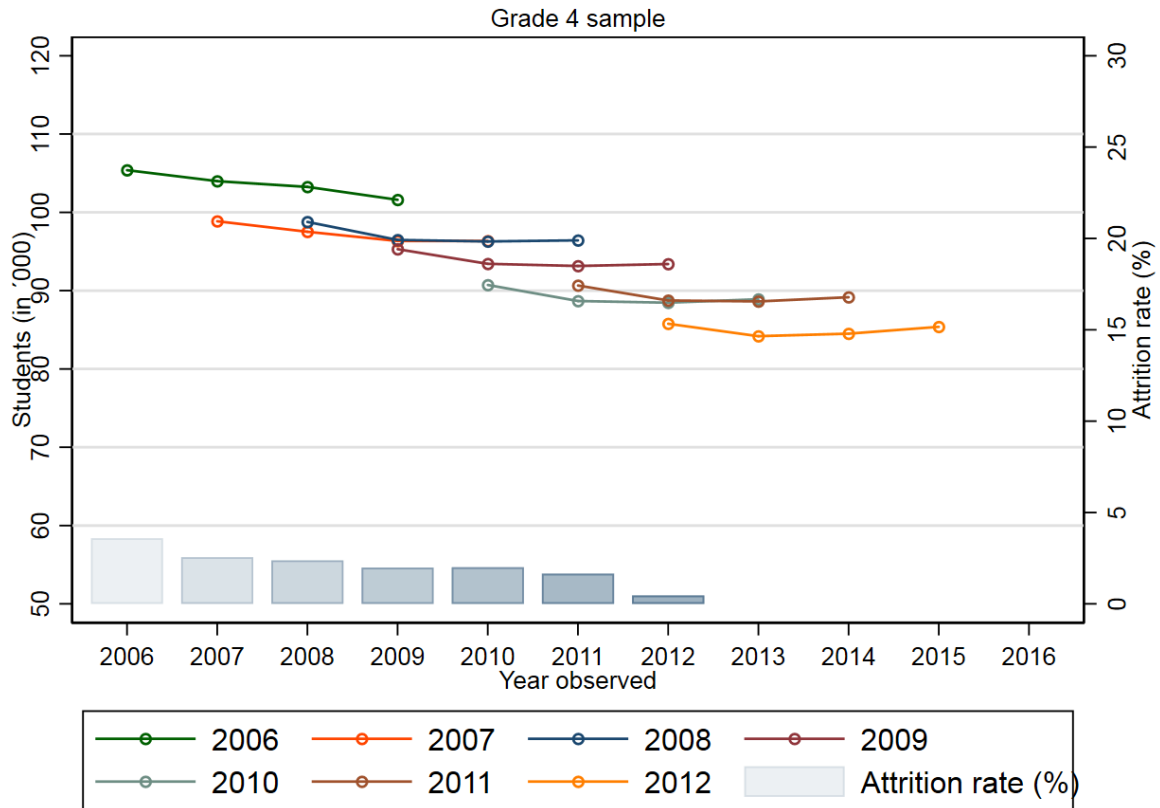
To construct the proxy variable for the household level of education we minimize problems with missing values by including the maximum level of education (in terms of years of study required) in the household (from either the father, the mother, or the legal guardian of the child, in case information for any of the parents is not available). All these characteristics are measured as of the school year in which the student is first enrolled in Grade 1.

In the Figures 8, 9 and 10, below, we document the rates of attrition in our analysis datasets, for each of the three samples and years at which the students started school. For longer panels (such as in Grades 6 and 9 samples) we tend to follow less students. In the administrative data, the track of a student is lost when she moves abroad, drops from the education system altogether, or dies. We may also lose track of students if these move to a different public or private school and the matching algorithm is unable to correctly assign the unique identifier to new instances of the same student in the system.

The main outcome variables in our regressions is constructed from the score points of students in Grades 4, 6 and 9 Math and Language national exams, as a proxy for student cognitive ability. Students sit – or sat – national exams or standardized achievement tests at the end of each of these grades. To the best of our knowledge, this is the only reliable assessment of cognitive ability that was systematically performed to a large cohort of children in Portugal. Exams are anonymized and scored by randomly selected evaluating teachers, which are not teachers of the students whose achievement is being scored.

The major advantage is that these tests were sat by the universe of eligible children by Grades 4, 6 and 9. Due to policy changes, Grade 4 exams were discontinued from the beginning of the 2015/16 school year onwards. Additionally, the score scale at which ability was measured has changed from school year 2012/2013 onwards, with the previous discrete scale (0-5) being insufficient to retain relevant variation across students. Given

FIGURE 8. PANEL ATTRITION RATES AND NUMBER OF TRACKED STUDENTS BY ENTRY COHORT (GRADE 4 SAMPLE, LATE DATASET)



Notes: Figure is based on the count of students in each cohort-grade cell for the sample of grade 4 students (i.e., for the sample in which students are followed from grade 1 to grade 4). The horizontal axis displays the year at which students are first observed in grade 1. The left-hand axis (associated with the lines in the graph) displays the number of students in thousands in a given cohort-grade cell. The right-hand side axis (associated with the bars in the graph) provides the attrition rate, i.e., the proportion of the number of students observed in grade 1 that are observed in grade 4, by cohort.

the constraints, we retain three years of observations of student outcomes (2013-2015) for Grade 4, as measured by exam scores (0-100 scale) by the end of the school year.

A simple analysis of the discrete distribution of exam scores shows significant manipulation of exam scores. The kinks in the distribution (at 20, 50, 75, and 90) coincide with threshold grade levels. The internal grades of primary education students are measured on a scale from 0-5 (from now on described as levels). These correspond to the threshold grade levels where kinks can be observed. Level 2, whose threshold is surpassed at 50 in the score scale is the most relevant here and the one where the kinks are most prominent. Exam scores below 50 are considered a fail, and students get a level 2 – a “negative” level, as it can lead to retention in the same grade. The raw distribution of exam scores strongly suggests that graders tend to upgrade scores that fall within a region of 5 points

FIGURE 9. PANEL ATTRITION RATES AND NUMBER OF TRACKED STUDENTS BY ENTRY COHORT (GRADE 6 SAMPLE, LATE DATASET)

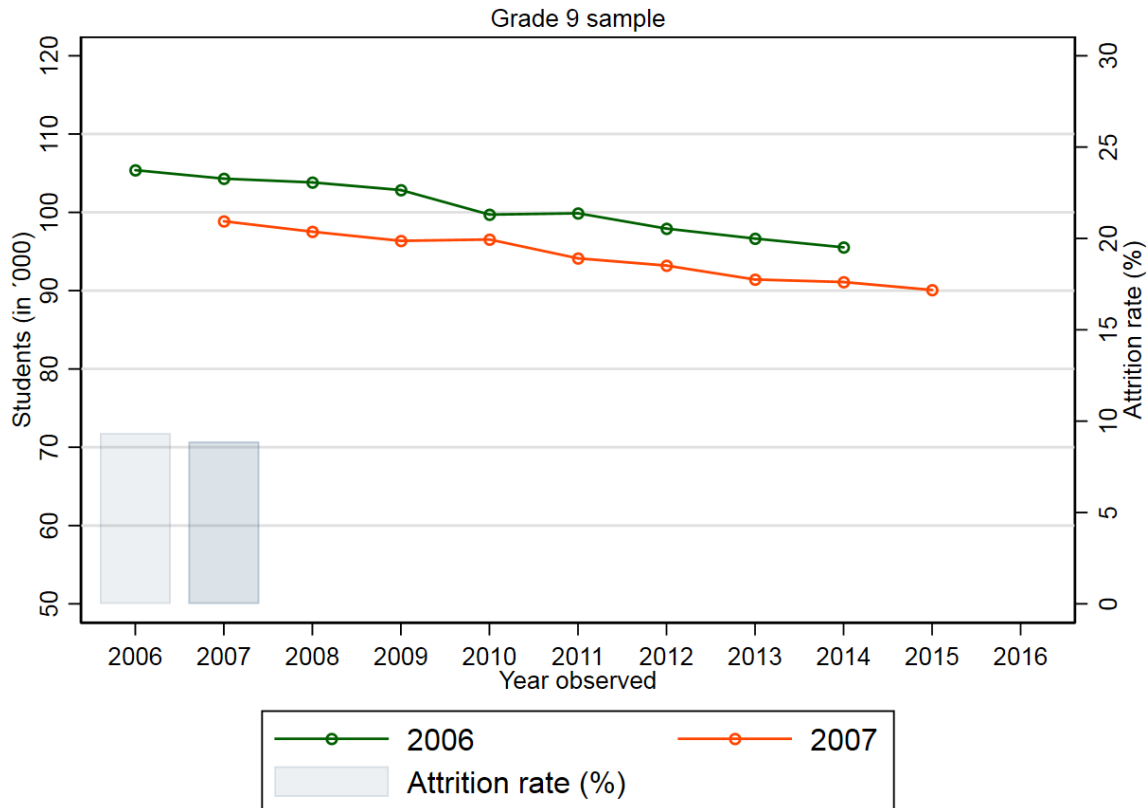


Notes: Figure is based on the count of students in each cohort-grade cell for the sample of grade 6 students (i.e., for the sample in which students are followed from grade 1 to grade 6). The horizontal axis displays the year at which students are first observed in grade 1. The left-hand axis (associated with the lines in the graph) displays the number of students in thousands in a given cohort-grade cell. The right-hand side axis (associated with the bars in the graph) provides the attrition rate, i.e., the proportion of the number of students observed in grade 1 that are observed in grade 6, by cohort.

below the relevant level threshold. For instance, most students that would have 49 points as score are upgraded to 50. Such evidence of manipulation is visible in intervals of 5 points.

While this sort of manipulation may be beneficial to students at the margin of threshold levels, it can significantly bias the analysis if one is to take unit changes in the exam score as informative of cognitive differences between individuals. In order to circumvent this concern, we collapse the 0-100 scale into a 0-20 scale. Such a scale has a couple of advantages. First, it still retains informative variation across students. Second, it almost entirely eliminates score manipulation bias and reduces noise. In the new scale, 5 – instead of 1 – underlying exam points are now considered informative of student abilities differences. In other words, students that have 47, 48, 49, or 50 are considered within the

FIGURE 10. PANEL ATTRITION RATES AND NUMBER OF TRACKED STUDENTS BY ENTRY COHORT (GRADE 9 SAMPLE, LATE DATASET)



Notes: Figure is based on the count of students in each cohort-grade cell for the sample of grade 9 students (i.e., for the sample in which students are followed from grade 1 to grade 9). The horizontal axis displays the year at which students are first observed in grade 1. The left-hand axis (associated with the lines in the graph) displays the number of students in thousands in a given cohort-grade cell. The right-hand side axis (associated with the bars in the graph) provides the attrition rate, i.e., the proportion of the number of students observed in grade 1 that are observed in grade 9, by cohort.

same range of cognitive ability. The distribution of scores is thus smoother by being less prone to bias introduced by score manipulation in the original scale.

National exams in Portugal are not standardized across years, making inter-temporal comparisons biased. In order to control for the difficulty of the exam in each year, we standardize the scores in the new scale (0-20) by subtracting the mean and dividing by the standard deviation of subject scores in each exam year and by grade. The outcome variable is thus interpretable in standard deviation units of the exam scores in a 0-20 scale.

Course- and track-specific exams include the Grade 11 subject exams of Physics, Biology, Geometry, Economics, Philosophy and Geography, as well as Grade 12 subject exams of Language, Math A, Physics, Biology, History, Geometry, Drawing and Economics.

B. Supplemental Tables

TABLE 5. IMPACT OF SCHOOL STARTING AGE ON STUDENT CHARACTERISTICS - COMPLIERS

Outcome:	Female		Immigrant		School social support		Unemployed dad		Computer at home		Higer education in HH	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
<i>Grade 4</i>												
School starting age	-0.016 (0.022)	-0.016 (0.022)	0.015 (0.008)	0.015 (0.008)	0.059 (0.020)	0.057 (0.018)	-0.008 (0.013)	-0.012 (0.013)	-0.014 (0.016)	-0.006 (0.015)	-0.007 (0.018)	0.007 (0.017)
Observations	72807	72807	72807	72807	72807	72807	72807	72807	72807	72807	72807	72807
<i>Grade 6</i>												
School starting age	0.004 (0.017)	0.005 (0.017)	0.007 (0.006)	0.007 (0.006)	-0.012 (0.011)	-0.014 (0.011)	0.002 (0.005)	0.003 (0.005)	0.004 (0.021)	0.005 (0.020)	-0.008 (0.014)	-0.010 (0.013)
Observations	94573	94573	94573	94573	94573	94573	94573	94573	94573	94573	94573	94573
<i>Grade 9</i>												
School starting age	0.019 (0.022)	0.020 (0.022)	0.008 (0.008)	0.008 (0.008)	-0.024 (0.012)	-0.027 (0.011)	0.009 (0.010)	0.010 (0.010)	0.018 (0.019)	0.019 (0.019)	-0.009 (0.012)	-0.015 (0.011)
Observations	59345	59345	59345	59345	59345	59345	59345	59345	59345	59345	59345	59345
Student controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Cohort Fes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: All coefficients are estimates of local-quadratic regressions. The excluded instrument is the post-cutoff indicator (τ). Both first and second stage regressions include a piecewise linear or quadratic function of birth dates (B) interacted with τ , depending on the specification indicated. A triangular kernel with a fixed 60-days bandwidth is used. Where indicated, regressions also control for cohort fixed effects as well as other individual covariates in the form of indicator variables for gender (1 if female), immigrant status (1 if first generation immigrant), reciprocity of school social support (1 if receiver), dad's unemployment status (1 if unemployed), access to computer at home (1 if yes) or if there is at least one child guardian with higher education (1 if yes). Robust standard errors clustered at the birthday level are presented in parentheses.

TABLE 6. IMPACT OF SCHOOL STARTING AGE ON STUDENT PERFORMANCE FOR NON-REPEATERS

Outcome: student performance at	Math				Language			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Grade 4</i>								
School starting age	0.251 (0.029)	0.280 (0.025)	0.245 (0.034)	0.275 (0.029)	0.342 (0.029)	0.370 (0.022)	0.325 (0.051)	0.356 (0.045)
Observations	31138	30064	45349	44192	63021	78135	57492	60766
Bandwidth (in days)	29	28	41	41	57	72	53	56
<i>Grade 6</i>								
School starting age	0.195 (0.031)	0.218 (0.024)	0.179 (0.035)	0.206 (0.026)	0.235 (0.029)	0.255 (0.025)	0.242 (0.031)	0.261 (0.027)
Observations	58094	56681	67816	73176	59572	60964	86475	86475
Bandwidth (in days)	43	42	49	54	43	45	64	64
<i>Grade 9</i>								
School starting age	0.073 (0.035)	0.105 (0.032)	0.150 (0.045)	0.152 (0.039)	0.170 (0.032)	0.182 (0.030)	0.180 (0.034)	0.187 (0.031)
Observations	38128	32081	32991	36401	34910	40944	51910	71246
Bandwidth (in days)	44	38	38	43	41	47	61	83
Polynomial order	Linear	Linear	Quadratic	Quadratic	Linear	Linear	Quadratic	Quadratic
Student controls	NO	YES	NO	YES	NO	YES	NO	YES
Cohort Fes	NO	YES	YES	YES	NO	YES	YES	YES

Notes: All coefficients are estimates of local regressions. The excluded instrument is the post-cutoff indicator (τ). Both first and second stage regressions include a piecewise linear or quadratic function of birth dates (B) interacted with τ , depending on the specification indicated. A triangular kernel with a data-driven MSE-optimal bandwidth is used. Where indicated, regressions also control for cohort fixed effects as well as individual covariates in the form of indicator variables for gender (1 if female), immigrant status (1 if first generation immigrant), reciprocity of school social support (1 if receiver), dad's unemployment status (1 if unemployed), access to computer at home (1 if yes), and fine-grained descriptions of the maximum level of education taken by the guardians of the child (e.g. primary education, lower secondary, bachelor degree, etc). Robust standard errors clustered at the birthday level are presented in parentheses.

TABLE 7. IMPACT OF SCHOOL STARTING AGE ON STUDENT OUTCOMES INCLUDING CONTROLS FOR DAY OF THE WEEK - LATE

Outcome:	Math performance			Language performance			Grade retention		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>(Until) Grade 4</i>									
School starting age	0.288 (0.028)	0.272 (0.024)	0.274 (0.024)	0.393 (0.075)	0.356 (0.048)	0.373 (0.035)	-0.041 (0.017)	-0.052 (0.013)	-0.053 (0.011)
Observations	35742	72001	61461	35767	72040	71283	35826	72155	68302
Left bandwidth (in days)	30	60	53	30	60	66	30	60	54
Right bandwidth (in days)	30	60	47	30	60	50	30	60	64
<i>(Until) Grade 6</i>									
School starting age	0.194 (0.042)	0.239 (0.030)	0.247 (0.028)	0.326 (0.031)	0.302 (0.031)	0.291 (0.024)	-0.071 (0.014)	-0.079 (0.013)	-0.078 (0.011)
Observations	46958	94251	87299	47043	94393	123601	47125	94573	101436
Left bandwidth (in days)	30	60	63	30	60	73	30	60	57
Right bandwidth (in days)	30	60	48	30	60	82	30	60	69
<i>(Until) Grade 9</i>									
School starting age	0.168 (0.060)	0.155 (0.038)	0.153 (0.037)	0.152 (0.049)	0.184 (0.036)	0.175 (0.031)	-0.032 (0.013)	-0.036 (0.011)	-0.039 (0.010)
Observations	29245	58995	49042	29407	59290	65373	29428	59345	54603
Left bandwidth (in days)	30	60	42	30	60	70	30	60	49
Right bandwidth (in days)	30	60	56	30	60	61	30	60	61

Notes: ll coefficients are estimates of local regressions. The excluded instrument is the post-cutoff indicator (τ). Both first and second stage regressions include a piecewise linear or quadratic function of birth dates (B) interacted with τ depending on the specification indicated. A triangular kernel with an data-driven MSE-optimal bandwidth is used. All regressions also control for cohort fixed effects, day of the week in which the student is born, as well as individual covariates in the form of indicator variables for gender (1 if female), immigrant status (1 if first generation immigrant), reciprocity of school social support (1 if receiver), dad's unemployment status (1 if unemployed), access to computer at home (1 if yes), and fine-grained descriptions of the maximum level of education taken by the guardians of the child (e.g. primary education, lower secondary, bachelor degree, etc). Robust standard errors clustered at the birthday level are presented in parentheses.

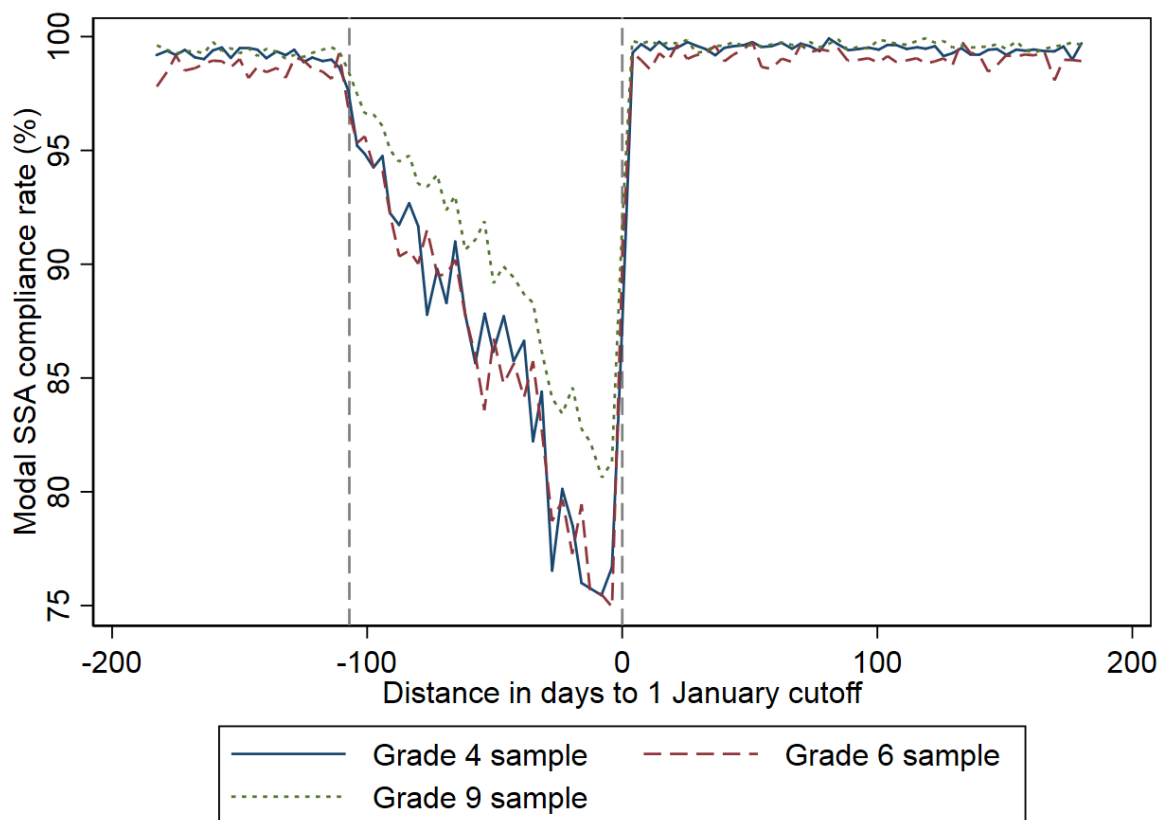
TABLE 8. SENSITIVITY ANALYSIS: IMPACT OF SCHOOL STARTING AGE ON STUDENT OUTCOMES FOR ALTERNATIVE BANDWIDTHS

Outcome:	Math performance		Language performance		Grade retention	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>(Until) Grade 4</i>						
School starting age	0.289 (0.022)	0.274 (0.024)	0.364 (0.034)	0.360 (0.040)	-0.060 (0.009)	-0.054 (0.011)
Observations	34866	53287	41015	65624	57032	66960
Bandwidth (in days)	29	44	34	53	47	54
<i>(Until) Grade 6</i>						
School starting age	0.262 (0.025)	0.255 (0.028)	0.294 (0.023)	0.291 (0.024)	-0.081 (0.010)	-0.079 (0.011)
Observations	64417	100530	61354	115385	61467	93673
Bandwidth (in days)	41	64	38	72	39	58
<i>(Until) Grade 9</i>						
School starting age	0.095 (0.024)	0.150 (0.038)	0.192 (0.024)	0.179 (0.032)	-0.047 (0.009)	-0.048 (0.010)
Observations	56529	39324	66312	69474	43488	76707
Bandwidth (in days)	57	39	66	70	43	76
Polynomial order	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic

Notes: All coefficients are estimates of local-quadratic regressions. The excluded instrument is the post-cutoff indicator (τ). Both first and second stage regressions include a piecewise quadratic function of birth dates (B) interacted with τ , depending on the specification indicated. As indicated, a triangular kernel with a 30-days, 60-days and a data-driven MSE-optimal choice of bandwidth, allowing bandwidths before and after the cutoff to differ, are used. All regressions also control for cohort fixed effects as well as individual covariates in the form of indicator variables for gender (1 if female), immigrant status (1 if first generation immigrant), reciprocity of school social support (1 if receiver), dad's unemployment status (1 if unemployed), access to computer at home (1 if yes), and fine-grained descriptions of the maximum level of education taken by the guardians of the child (e.g. primary education, lower secondary, bachelor degree, etc). Robust standard errors clustered at the birthday level are presented in parentheses.

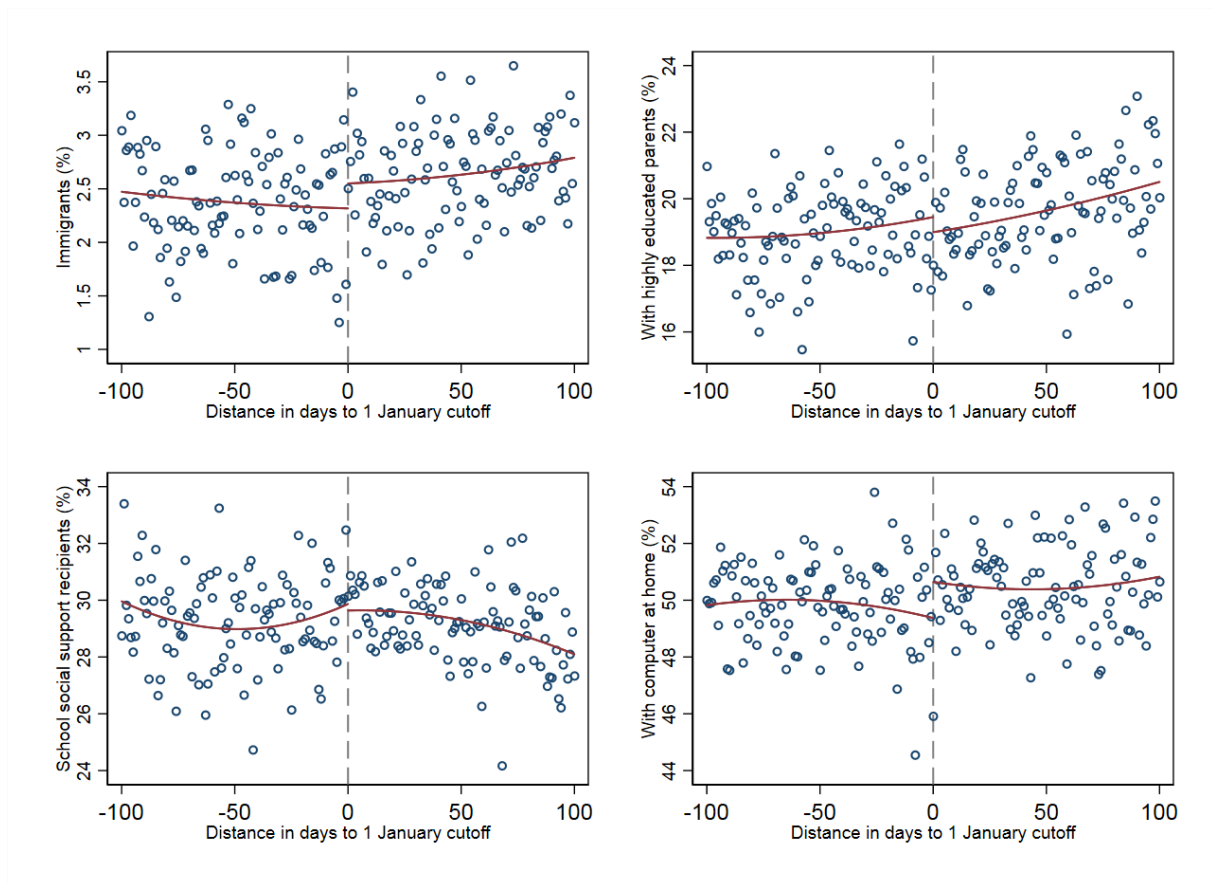
C. Supplemental Figures

FIGURE 11. COMPLIANCE RATES WITH THE 1 JANUARY CUTOFF BY GRADE SAMPLE



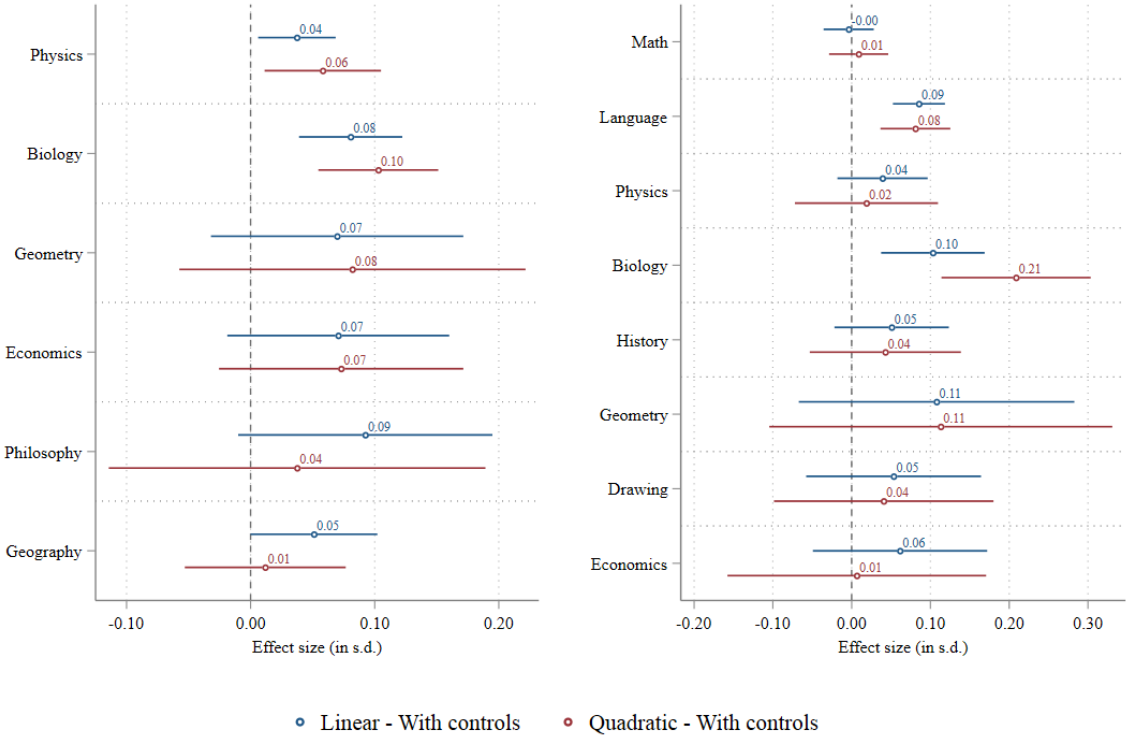
Notes: Figure is based on cohorts of students that entered a regular curriculum program in a public school in continental Portugal between 2007 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in Grade 1. Lines represent the trend (by day of birth) in compliance rates, i.e., the ratio of students that did not differ school entrance to the next school year given the modal school starting age for students born in a given day and the number of students born in each birth date bins, by grade sample. The first dashed line represents the 16 September cutoff, while the second dashed line represents the 1 January cutoff.

FIGURE 12. STUDENT CHARACTERISTICS BY BIRTHDATE CLOSE TO THE CUTOFF



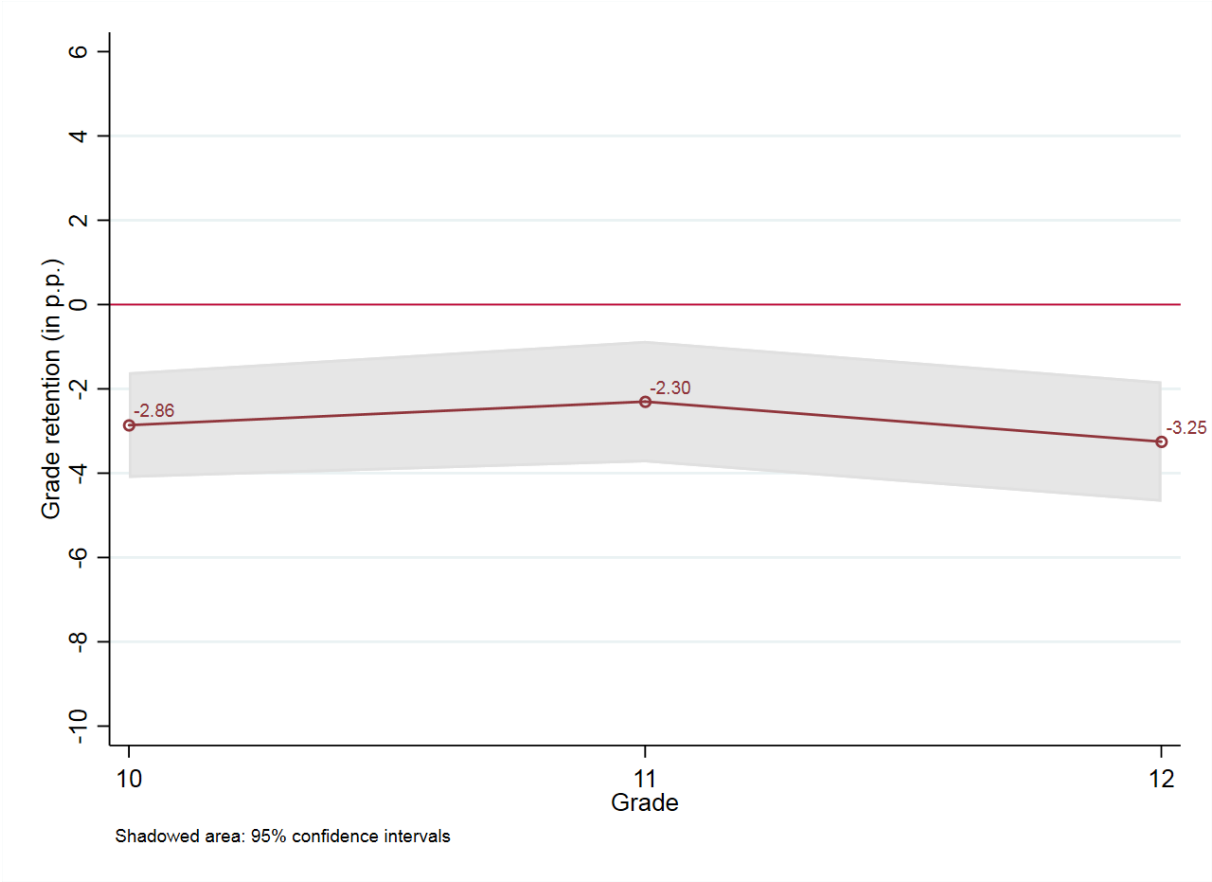
Notes: Figure is based on cohorts of students that entered a public school in continental Portugal between 2007 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in Grade 1. Horizontal axis represents the birth date relative to the cutoff date of 1 January. Each hollow circle represents within birth day cell averages of immigrant status (top left panel), students with highly educated parents (top right panel), school support recipients (bottom left panel) and students with access to a computer at home (bottom right panel). Solid lines represent fitted values from a piecewise quadratic spline. Vertical dashed lines identify 1 January (= 0).

FIGURE 13. IMPACT OF DELAYED ENTRANCE ELIGIBILITY ON STUDENT ACHIEVEMENT IN GRADES 11 AND 12 NATIONAL EXAMS BY SUBJECT - ITT



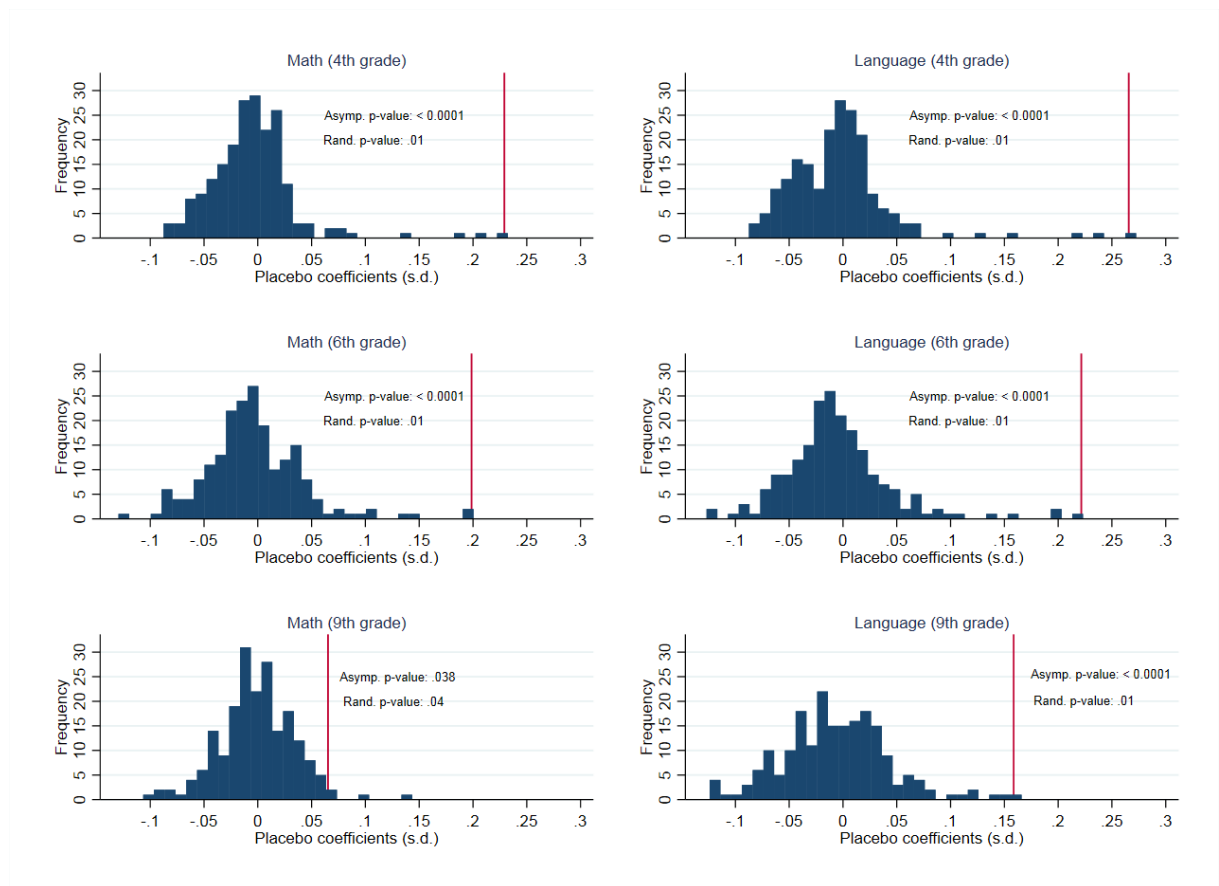
Notes: The left-hand panel refers to exams in grade 11. The right-hand side refers to exams in grade 12. Point estimates are coefficients of local regressions. The regressor of interest is the post-cutoff indicator (τ). All regressions include a piecewise quadratic function of birth dates (B) interacted with τ . A triangular kernel with data-driven optimal bandwidths is used. Regressions also control for cohort fixed effects as well as all other individual covariates. The bars 95% confidence intervals for clustered standard errors at the birthday-level.

FIGURE 14. IMPACT OF SCHOOL STARTING AGE ON GRADE RETENTION THROUGHOUT HIGH SCHOOL - ITT



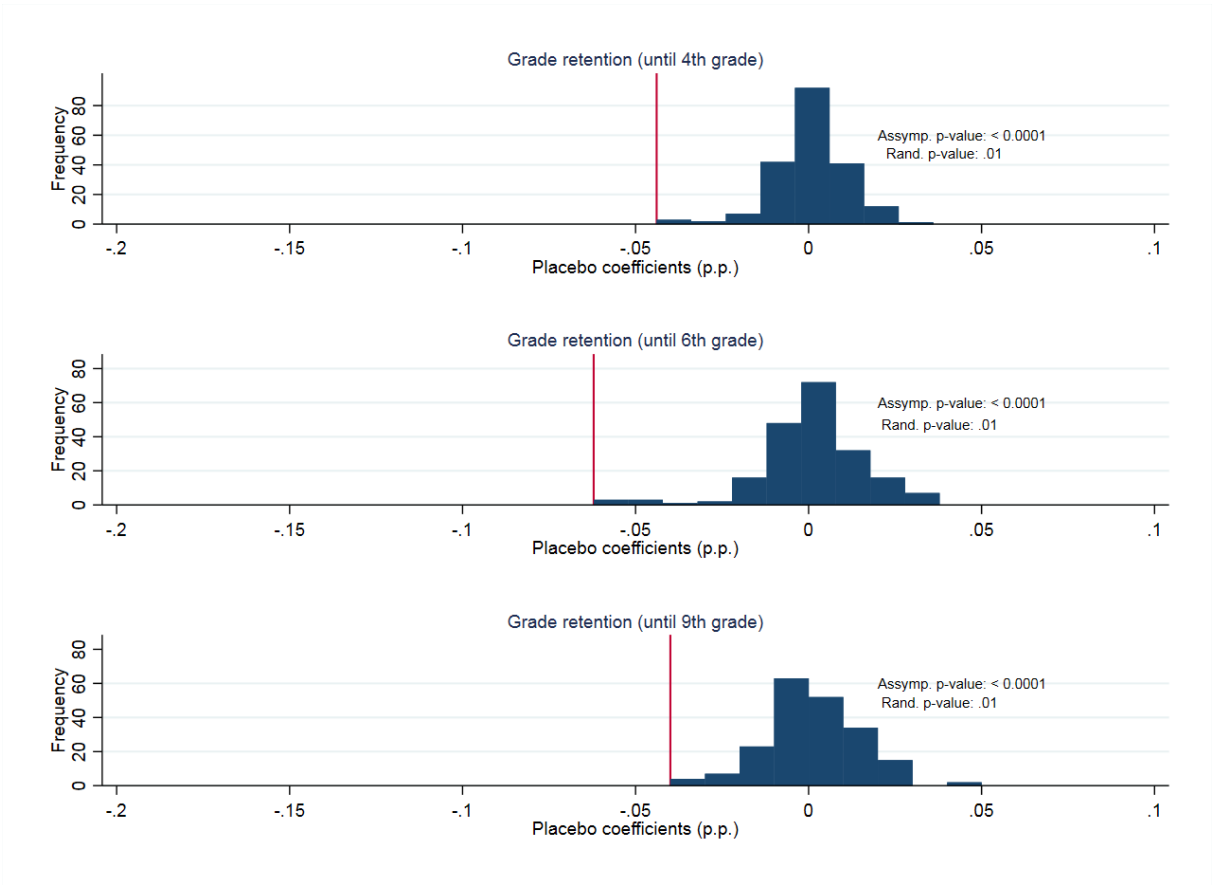
Notes: Point estimates are coefficients of local regressions. The regressor of interest is the post-cutoff indicator (τ). All regressions include a piecewise quadratic function of birth dates (B) interacted with τ . A triangular kernel with data-driven optimal bandwidths is used. Regressions also control for cohort fixed effects as well as all other individual covariates. The shaded area represents 95% confidence intervals for clustered standard errors at the birthday-level.

FIGURE 15. DISTRIBUTIONS OF PLACEBO AND TRUE INTENT-TO-TREAT EFFECTS ON ACHIEVEMENT BY SUBJECT AND GRADE



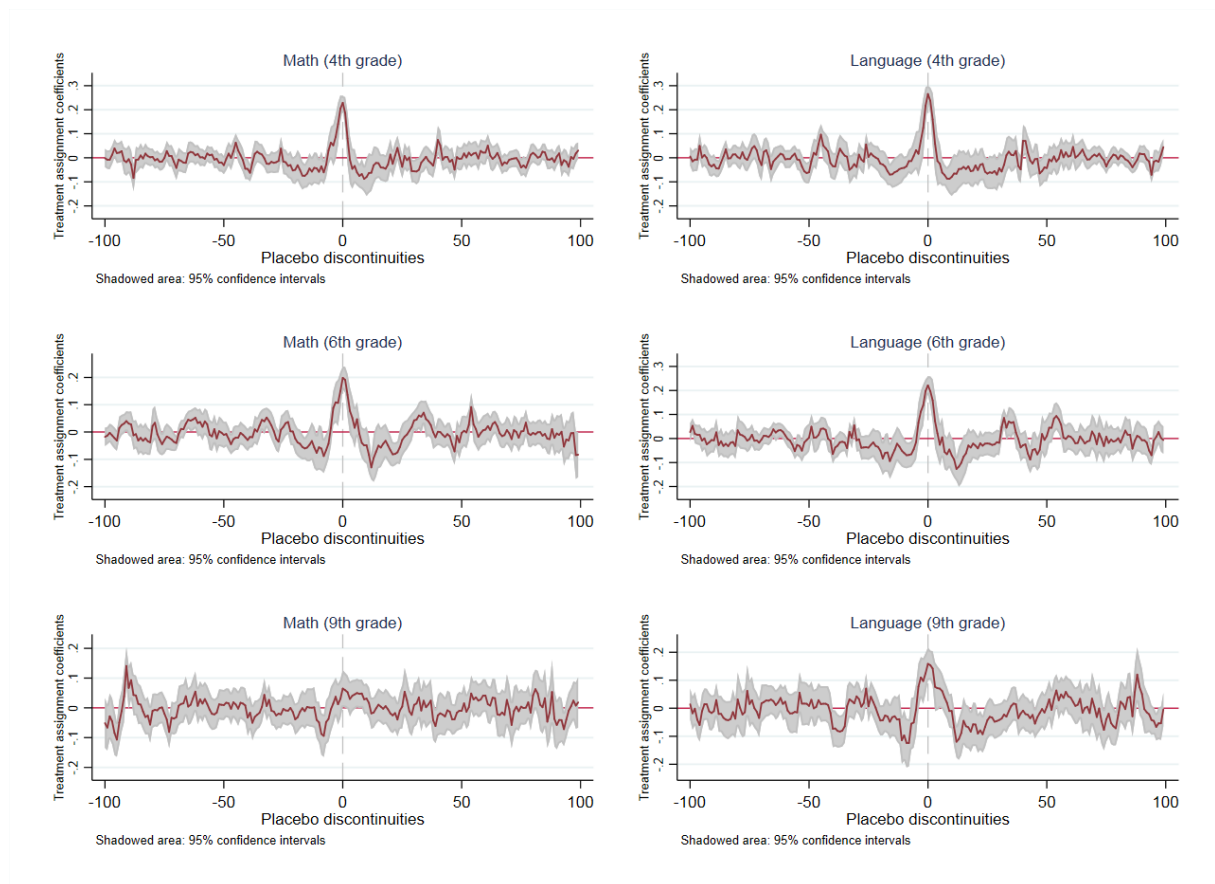
Notes: Figure is based on cohorts of students that entered a regular curriculum program in a public school in continental Portugal between 2007 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in Grade 1. Each panel is an histogram of estimated intent-to-treat effects for a total of 200 potential cutoffs (including the true one). Cutoffs considered are all where $B_i = \{100, 99\}$. For each placebo cutoff the specification estimated is the one in Column 1 (for Math) and Column 2 (for Language) in Table 3, with varying MSE-optimal bandwidths. The vertical red line represents the position of the coefficient of the true cutoff at the distribution of placebo cutoffs. Randomization-based p-values, computed through the software package `rdpermute` and according to (Ganong and Jäger, 2018), are presented under the asymptotic p-values for the preferred specification at the true cutoff.

FIGURE 16. DISTRIBUTIONS OF PLACEBO AND TRUE INTENT-TO-TREAT EFFECTS ON GRADE RETENTION



Notes: Figure is based on cohorts of students that entered a regular curriculum program in a public school in continental Portugal between 2007 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in Grade 1. Each panel is an histogram of estimated intent-to-treat effects for a total of 200 potential cutoffs (including the true one). Cutoffs considered are all where $B_i = \{100, 99\}$. For each placebo cutoff the specification estimated is the one in Column 1 in Table 4, with varying MSE-optimal bandwidths. The vertical red line represents the position of the coefficient of the true cutoff at the distribution of placebo cutoffs. Randomization-based p-values, computed through the software package `rdpermut` and according to (Ganong and Jäger, 2018), are presented under the asymptotic p-values for the preferred specification at the true cutoff.

FIGURE 17. INTENT-TO-TREAT EFFECTS BY POTENTIAL CUTOFF, SUBJECT AND GRADE



Notes: Figure is based on cohorts of students that entered a regular curriculum program in a public school in continental Portugal between 2007 and 2013, and were at least 5 years-old and at most 8 years-old when first enrolled in Grade 1. Each panel shows estimated intent-to-treat effects for a total of 200 potential cutoffs (including the true one) along the y-axis. Cutoffs considered are all where $B_i = \{100, 99\}$ spread along the x-axis. For each placebo cutoff the specification estimated is the one in Column 1 (for Math) and Column 2 (for Language) in Table 3, with varying MSE-optimal bandwidths. The vertical dashed line represents the position of the coefficient of the true cutoff at the distribution of placebo cutoffs.

Nova School of Business & Economics

Campus de Carcavelos

Rua da Holanda 1

2775-405 Carcavelos | Portugal

novasbe.pt

Accredited by



Member of

