



NATIONAL  
CENTER *for* ANALYSIS of LONGITUDINAL DATA *in* EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

*A program of research by the American Institutes for Research with Duke University, Northwestern University, Stanford University, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington*



## School Starting Age and Cognitive Development

Elizabeth Dhuey  
David Figlio  
Krzysztof Karbownik  
Jeffrey Roth

---

# School Starting Age and Cognitive Development

Elizabeth Dhuey  
*University of Toronto*

David Figlio  
*Northwestern University*

Krzysztof Karbownik  
*Northwestern University*

Jeffrey Roth  
*University of Florida*

## Contents

---

Acknowledgements.....	ii
Abstract.....	iii
1. Introduction.....	1
2. Estimation.....	4
3. Results.....	9
4. Conclusions.....	17
References.....	19
Tables & Figures.....	23
Appendix.....	39

## Acknowledgements

---

We are grateful to the Florida Departments of Education and Health for providing the de-identified, matched data used in this analysis. Figlio and Roth appreciate funding from the U.S. Department of Education, and Figlio appreciates funding from the National Institutes of Health and the Bill and Melinda Gates Foundation. We appreciate helpful feedback from Todd Elder, Jennifer Heissel, Umut Özek and Helena Skyt Nielsen and conference participants at Ce2 Workshop. The conclusions expressed in this paper are those of the authors and do not represent the positions of the Florida Departments of Education and Health or those of our funders.

This research was supported by the National Center for the Analysis of Longitudinal Data in Education Research (CALDER), which is funded by a consortium of foundations. For more information about CALDER funders, see [www.caldercenter.org/about-calder](http://www.caldercenter.org/about-calder). All opinions expressed in this paper are those of the authors and do not necessarily reflect the views of our funders or the institutions to which the authors are affiliated.

CALDER working papers have not undergone final formal review and should be cited as working papers. They are intended to encourage discussion and suggestions for revision before final publication. Any opinions, findings, and conclusions expressed in these papers are those of the authors and do not necessarily reflect the views of our funders.

CALDER • American Institutes for Research  
1000 Thomas Jefferson Street N.W., Washington, D.C. 20007  
202-403-5796 • [www.caldercenter.org](http://www.caldercenter.org)

## **School Starting Age and Cognitive Development**

Elizabeth Dhuey, David Figlio, Krzysztof Karbownik, Jeffrey Roth

CALDER Working Paper No. 191

June 2018

### **Abstract**

We present evidence of a positive relationship between school starting age and children’s cognitive development from age 6 to 18 using a fuzzy regression discontinuity design and large-scale population-level birth and school data from the state of Florida. We estimate effects of being old for grade (being born in September versus August) that are remarkably stable – always around 0.2 SD difference in test scores – across a wide range of heterogeneous groups, based on maternal education, poverty at birth, race/ethnicity, birth weight, gestational age, and school quality. While the September-August difference in kindergarten readiness is dramatically different by subgroup, by the time students take their first exams, the heterogeneity in estimated effects on test scores effectively disappears. We do, however, find significant heterogeneity in other outcome measures such as disability status and middle and high school course selections. We also document substantial variation in compensatory behaviors targeted towards young for grade children. While the more affluent families tend to redshirt their children, young for grade children from less affluent families are more likely to be retained in grades prior to testing. School district practices regarding retention and redshirting are correlated with improved outcomes for the groups less likely to use those remediation approaches (i.e., retention in the case of more-affluent families and redshirting in the case of less-affluent families.) Finally, we find that very few school policies or practices mitigate the test score advantage of September born children.

Keywords: school starting age, educational attainment, socioeconomic gradient, redshirting, grade retention

# 1 Introduction

One of the largest questions that looms in a parent’s mind while thinking about enrolling their children in primary school for the first time is whether or not they are “ready” for school. This question has been made more fraught as the popular media frequently reports on research findings regarding the negative effects of entering school too young (e.g. [Weil 2007](#)). In response, an increasing number of parents in the United States have been delaying sending their children to kindergarten because they believe doing so will give them an advantage over their peers, whether academically, socially, or even athletically ([Deming and Dynarski 2008](#)). This practice is called redshirting. As an alternative, schools can retain children in early grades in order to allow them to mature enough for primary school challenges. Despite an ever growing academic and popular culture literature, however, it is still unclear what disadvantage certain children face due to their age at school entry and what the best remediation method is for that disadvantage.

The age distribution at school entry exists because most states in the United States and jurisdictions worldwide have a single specific cutoff date which determines when a student can enter primary school. For example, in Florida, a child is eligible to enter kindergarten if s/he turns five years old by September 1st of the relevant school year. These cutoffs effectively cause the oldest child to be up to one year older than the youngest child in a school cohort. A number of recent studies have found that children who enter school at an older age than their classmates have a variety of short- and medium-run advantages such as scoring higher on standardized exams through primary and secondary school<sup>1</sup>, having higher development of non-cognitive skills ([Lubotsky and Kaestner 2016](#)), and being less likely to commit a crime ([Cook and Kang 2016](#); [Depew and Eren 2016](#)). Some other examples of outcomes investigated in this literature include high school leadership ([Dhuey and Lipscomb 2008](#)), becoming a corporate CEO ([Du et al. 2012](#)) or politician ([Muller and Page 2016](#)), secondary school track placement ([Bedard and Dhuey 2006](#); [Puhani and Weber 2007](#); [Muhlenweg and Puhani 2010](#); [Schneeweis and Zweimuller 2014](#)), fertility ([Black et al. 2011](#); [McCrary and Royer 2011](#); [Tan 2017](#); [Pena 2017](#)), and disability identification, mental health and special education service uptake.<sup>2</sup> All these findings together suggest that early differences in maturity can propagate through the human capital accumulation process into later life and may have important implications for adult outcomes and productivity. At the same time, the evidence regarding the relationship between being older at school entry and a variety of adult outcomes is more mixed. Previous research includes inconclusive results on both academic attainment<sup>3</sup> and wages.<sup>4</sup>

We use detailed population-level administrative data from the state of Florida, where we observe matched birth and schooling outcomes, to study the effect of age at school entry. In doing so, we make three principal contributions to the literature on the effects of school starting age. First, we offer the most comprehensive set of controls for potential selection into timing of birth yet considered in the literature, and bring together in the same research design the two most compelling

---

<sup>1</sup>See for example: [Bedard and Dhuey \(2006\)](#); [Datar \(2006\)](#); [Crawford et al. \(2007\)](#); [Puhani and Weber \(2007\)](#); [McEwan and Shapiro \(2008\)](#); [Elder and Lubotsky \(2009\)](#); [Smith \(2009\)](#); [Crawford et al. \(2010\)](#); [Sprietsma \(2010\)](#); [Kawaguchi \(2011\)](#); [Robertson \(2011\)](#); [Nam \(2014\)](#); [Lubotsky and Kaestner \(2016\)](#); [McAdams \(2016\)](#); [Landerso et al. \(2017b\)](#); and [Attar and Cohen-Zada \(2017\)](#).

<sup>2</sup>[Black et al. \(2011\)](#); [Dhuey and Lipscomb \(2010\)](#); [Elder \(2010\)](#); [Elder and Lubotsky \(2009\)](#); [Evans et al. \(2010\)](#); [Morrow et al. \(2012\)](#); and [Dee and Sievertsen \(2017\)](#)

<sup>3</sup>[Dobkin and Ferreira \(2010\)](#) and [Black et al. \(2011\)](#) find little to no effect on academic attainment whereas [Bedard and Dhuey \(2006\)](#); [Kawaguchi \(2011\)](#); [Fredriksson and Ockert \(2014\)](#); [Cook and Kang \(2016\)](#); [Pena \(2017\)](#) find a positive effect of being older on academic attainment. However, [Hemelt and Rosen \(2016\)](#) and [Hurwitz et al. \(2015\)](#), find the opposite to be true.

<sup>4</sup>For instance, [Fredriksson and Ockert \(2014\)](#), [Kawaguchi \(2011\)](#), and [Pena \(2017\)](#) find that older children at school entry earn higher wages. In contrast, [Black et al. \(2011\)](#), [Dobkin and Ferreira \(2010\)](#), [Fertig and Kluve \(2005\)](#), [Nam \(2014\)](#) and [Larsen and Solli \(2017\)](#) find no such long-term wage effects.

approaches used in the literature to attempt to correct for this selection. Specifically, we present the first evidence from an environment in which we can execute a regression-discontinuity design, comparing children whose ages mean that they would “naturally” be the oldest in their class to those whose ages mean that they would “naturally” be the youngest in their class, while at the same time making this comparison *within families*. Comparing one child born in August to their sibling born in September dramatically reduces the likelihood that observed results are due to unobserved differences in families who time births for August versus those who time births for September.<sup>5</sup> Some studies (Cook and Kang 2016; Elder and Lubotsky 2009) have made use of the regression discontinuity approach before, and one study (Black et al. 2011) has made sibling comparisons, but we are the first to simultaneously compare siblings who just barely met or missed the threshold for school attendance in a given academic year. We also are able to control for conditions and treatments surrounding pregnancy and birth. We ultimately find that these extra controls do not alter our results, indicating that omitted-variables bias in the extant literature is likely not as large as some might fear *ex ante*. At the same time, since we can track students from birth to schooling we document the demographic differences between these two populations and find that the estimation sample is negatively selected. This issue may be very common in other data sets used in this literature, however, it is not possible to address it using only school records. Thus, we carry out a bounding exercise to determine the degree at which this might influence our results.

Our second contribution involves a comprehensive study of the heterogeneous effects of school starting age. Families differ dramatically in terms of the degree to which they actively attempt to remediate their children’s being young for grade. Schanzenbach and Howard (2017), for example, report that summer-born sons of college-educated parents are nearly four times as likely to be redshirted as are summer-born sons of high-school educated parents. Similarly, Cook and Kang (2018) document differences in redshirting in various groups in North Carolina. If families differ this remarkably regarding how they treat young children, it stands to reason that the effects of school starting age might be different for different groups of children. To date, however, there has been little comprehensive research examining the heterogeneous effects of school starting age in the US context, largely due to limitations in US administrative data, and the studies that exist have generally not been able to carry out the analysis using the preferred regression discontinuity approach or using exhaustive individual and family background information.<sup>6</sup> This paper represents the most robust analysis of heterogeneous effects of school starting age in a regression discontinuity framework. Moreover, we consider a wide range of cuts of the data on a wide range of outcomes (including test scores, disability and gifted status, middle and high school course selection and high school

---

<sup>5</sup>Of course, it’s always possible that a family might, for some reason, intentionally time one birth for September but not do so for another birth, but at least any characteristics of a family that are invariant across siblings will be absorbed in the family fixed effect.

<sup>6</sup>School registers in the US rarely contain background variables other than race/ethnicity and free lunch status, so only with either a match to birth certificates or the use of Census style data sets researchers can study heterogeneous effects with regard to a wide range of background factors. Heterogeneity has been investigated in settings with broader access to registry data, and thus background variables: Chile (McEwan and Shapiro (2008), who find little differences in the effects of school starting age by parental education); Denmark (Landerso et al. (2017b), who find evidence for smaller adverse effects of school starting age on crime for groups with both better educated mothers and unemployed fathers); Israel (Attar and Cohen-Zada (2017), who find little differences by parental education); Norway (Black et al. (2011), who find little differences by predicted family affluence); and Sweden (Fredriksson and Ockert (2014), who find larger advantage in both education and earnings for children of lower educated parents). In the U.S., Datar (2006) and Elder and Lubotsky (2009) estimate the effects of school starting age by family SES background but find conflicting results. Cook and Kang (2016) use population-level data and a regression discontinuity analysis, but because they focus on crime and delinquency they only investigate various definitions of significant disadvantage. Hemelt and Rosen (2016) examine longer run outcomes in a regression discontinuity framework by race/ethnicity and poverty proxy (FRL), however, they do not observe actual kindergarten entry.

graduation). We stratify by maternal education; by poverty at birth; by race and ethnicity; by birth weight; by gestational age; and by experienced school quality; as well as by gender interacted with many of these stratifications. These stratifications are potentially important because they illustrate how age effects might differ depending on generalized school factors or by biological factors. For example, we know that better neonatal health, as proxied by higher birth weight, has a positive effect on longer-run outcomes such as educational attainment, IQ, and life-cycle earnings (Black et al. 2007; Figlio et al. 2014; Bharadwaj et al. 2017). Therefore, it is natural to think that maybe birth weight might dynamically interact with a child’s age relative to their classmates within the human capital production function framework (Cunha et al. 2010). This complementarity could also occur to the degree to which educators have difficulty distinguishing between innate ability and maturity. Birth weight and its subsequent effect on childhood height and weight may make it difficult to disentangle maturity from ability as larger children may appear to be more mature due to their physical stature. Likewise, gestational age is another avenue one might suspect could affect the age gap (Figlio et al. 2016; Garfield et al. 2017). These interactions between initial birth endowments and school starting age have never been studied in the extant literature.

We find remarkable stability in the effects of school starting age on test scores across exceptionally different groups of people, and despite differences in both remediation strategies and non-test score outcomes like disability diagnoses or course enrollment. We further find that the August-September gap in test scores is not mediated by measured school quality. This pattern of results suggests that the academic remediation for being young for grade may be more challenging than those who seek to remediate might believe. In the non-test score outcomes, the August-September difference is smaller for higher educated and higher income families on being identified with a disability (in both the behavioral and cognitive domains) and taking advanced courses in middle and high school. Our heterogeneity estimates for high school graduation outcomes are not precise enough to infer any particular pattern.

The finding of an exceptional lack of heterogeneous effects of school starting age on test scores leads us to our third contribution. In this paper we directly explore the potential efficacy of school policies and attempted remediation techniques. First, we explore the interaction of school level policies with age at school entry. We are able to explore twenty different programs or policies and find that only three interact with the estimates for school starting age - the practice of block scheduling, summer school requirements for grade advancement among low-performing students and class size. Interestingly both the first two policies and larger class size increase the August-September difference.

Next we turn to a combination of parental and school remediation strategies. Like Schanzenbach and Howard (2017), we show in our population-level data that there exist substantial differences in remediating behaviors among parents of different socioeconomic groups, with higher-SES parents being more likely to redshirt their children than lower-SES parents. Conversely, children who are from lower-SES families are more likely than their higher-SES counterparts to be retained in early grades. As a potential consequence of these two sets of actions, by the time children reach third grade, the ratio of September- to August-born children who are below grade for age is roughly equal across SES groups. This pattern of behaviors could help to explain why we document such a strong SES gradient in the September-August difference in kindergarten readiness (where high-SES families are disproportionately likely to redshirt August-born children) but no SES gradient in the September-August difference in third grade test scores.

Armed with this evidence, we then turn to the following questions: Do school district practices related to redshirting and retention help remediate the relative age effect? And are remediation approaches like redshirting or grade retention more effective when used by groups for whom the approach is unusual? While we cannot obtain strong causal evidence on this point, we produce



suggestive evidence that indicates that this may be the case. Florida has large county-level school districts that vary dramatically in the rate of redshirting or retention of August-born children. Medium-to-large Florida school districts range in their August-born redshirting rates from fewer than two percent to over ten percent, and range in their August-born early-grade retention rates from 20 percent to 45 percent. Districts with relatively high redshirting rates have higher-than-usual redshirting rates for both low-SES and high-SES August-born children alike (the correlation between overall August redshirting rates and low-SES August redshirting rates in these districts is 0.737) and districts with relatively high early-grade retention rates have higher-than-usual early-grade retention rates for all SES groups (the correlation between overall August early-grade retention rates and high-SES August early-grade retention rates in these districts is 0.745). We find that districts where redshirting is more prevalent have lower August-September differences in test scores for low-SES families (for whom redshirting is less common), and that districts where early-grade retention is more prevalent have lower August-September differences in test scores for high-SES families (for whom early-grade retention is less common). These findings, while merely suggestive, indicate a potential role for strategically-deployed instructional policies and practices to help modify preparation differences caused by school starting age cutoffs.

## 2 Estimation

### 2.1 Data

We used birth records from the Florida Department of Health for all children born in Florida between 1992 and 2000, merged with school records maintained by Florida Department of Education for the academic years 1997-98 through 2012-13. The children were matched along four dimensions: first and last names, date of birth, and social security number. Rather than conducting probabilistic matching, the match was performed such that a child would be considered matched so long as (1) there were no more than two instances of modest inconsistencies, and (2) there were no other children who could plausibly be matched using the same criteria. Common variables excluded from the match were used as checks of match quality. These checks confirmed a very high and clean match rate. In the overall match on the entire population, the sex recorded on birth records disagreed with the sex recorded in school records in about one-one thousandth of one percent of cases, suggesting that these differences are likely due to typos in the birth or school records.

There were 1,220,803 singleton births with complete demographic information in Florida between 1994 and 2000, and of these 989,054 children were subsequently observed in Florida public schools data, representing an 81.0 percent match rate. The match rate is almost identical to the percentage of children who are born in Florida, reside there until schooling age, and attend public school, as computed using data from the decennial Census and American Community Survey for years 2000 through 2009 (Figlio et al. 2014). Multiple births are excluded from the analysis while siblings are identified in school districts representing the vast majority of Florida households. Figlio et al. (2014) discuss the differences between these school districts, which are disproportionately non-rural, and the state as a whole.

The data include a wide variety of demographic characteristics of the mother that are gathered from the Florida birth certificate. These include racial-ethnic information, education level, marital status at the time of the child's birth, and place of residence. We also have demographic characteristics of the father if he appears on the birth certificate, and health and demographic characteristics of the newborn. We observe birth weight, gestational age and indicators for any maternal health problems, whether or not they are related to the pregnancy. Finally, we know if the birth was paid for by Medicaid, an indicator of living in or near poverty at the time of the birth.

Moving to school records, we can observe school quality as defined by the state of Florida via its school accountability system. Since 1999, the Florida Department of Education has awarded each of its public schools a letter grade ranging from A (best) to F (worst). Initially, the grading system was based mainly on average proficiency rates on the FCAT standardized exam. Beginning in 2002, grades were based on a combination of average FCAT proficiency rates and average student level FCAT test score gains from year to year. We utilize this information to construct a time-invariant school quality measure. For each school, we compute a simple average of the observed gain scores between 2002 and 2013, as measured by the Florida Department of Education, which we then convert into a percentile rank in the observed gains distribution across Florida schools. These values are then attached to students for each school year and school they attend.

Our data also include information about school policies and practices that come from surveys administered to all public school principals in Florida. School surveys were conducted three times in school years: 1999-2000, 2001-2002, and 2003-2004 (Rouse et al. 2013). In our analysis, we use the first survey wave, which asked a broader set of questions, and we code schools as using a given policy if they responded “yes” to a question.<sup>7</sup> These questions and additional information are provided in Appendix A1. We use five questions and assign school answers to students attending grade one in a given school irrespective if they attended or not this school in a year when the survey was conducted.

We focus on a variety of short- and medium- term outcomes: kindergarten readiness, parental holding back behavior (redshirting), school retention behavior, test scores from grade three through eight, disability and gifted status, middle and high school course selection, as well as high school graduation. Kindergarten readiness is measured by a universally-administered screening at the entrance to kindergarten. The Florida Department of Education recorded readiness measures for those who entered kindergarten in fall 2001 and before, and those who entered kindergarten in fall 2006 or later.<sup>8</sup> Because of this data restriction we are unable to use this outcome for children born between 1997 and 1999.

Holding back or redshirting is defined as an indicator variable that equals to one if a child has higher than expected, based on date of birth, age at the time of first observation in school records in either kindergarten or grade one.<sup>9</sup> These are six or above for kindergarten and seven or above for grade one. We view redshirting as primarily a parental decision. School retention prior to grade three is defined as an indicator variable that equals to one if child is observed twice in the same grade. Florida has mandatory retention policy in grade three, and thus we are unable to utilize retention as school behavior measure after grade two (Schwerdt et al. 2015; Ozek 2015).

Our measure of academic performance is based on Florida Comprehensive Assessment Test (FCAT) in mathematics and reading, a state-wide standardized yearly assessment of all students in Florida conducted in grades three through ten. In this paper we focus on test scores in grades three through eight, because curriculum differences make interpersonal test score comparisons relatively difficult in high school (e.g., one tenth grader is taking algebra while another is enrolled in calculus). Therefore, each child in the sample can contribute up to six observations, one for each grade

---

<sup>7</sup>Our results are substantively unchanged if we use multiple survey waves and a more limited set of questions.

<sup>8</sup>In the early round of kindergarten readiness assessments, teachers administered a readiness checklist of academic and behavioral skills designed by the state Department of Education with a dichotomous ready/not-ready measure recorded in state records. In the later round of kindergarten readiness, the state universally implemented the DIBELS assessment aimed at measuring early pre-literacy skills. DIBELS is a discrete measure that we dichotomize using the approach described in Figlio et al. (2013) so that the percentage identified as kindergarten ready corresponds to the percentage in the later assessment. In our analysis sample, the birth cohorts which took the kindergarten readiness assessment are those born between 1994 and 1996 (kindergarten checklist) and those born in 2000 (DIBELS).

<sup>9</sup>Kindergarten attendance in Florida is not mandatory but it is heavily subsidized and 95.8 percent of children in school records whom we observe in grade one also attended kindergarten. In our estimation sample, this fraction is 89.9 percent.

observed. For brevity we average the math and reading test scores but we present main results split by reading and math in the Appendix Table A2. We also average the test scores across grades but the results for individual grades are presented in Figure 1.

Information on disability and gifted status comes from school records, and is based on mutually exclusive categories. A child may have multiple disabilities and we observe all of these but we focus our analysis on what is defined in the data as a primary exceptionality. We divide disabilities into three groups: cognitive, behavioral, and physical, and when we estimate the effects for one of the sub-types we always compare it to individuals without any disability.<sup>10</sup> Gifted status is defined by Florida Department of Education as “one who has superior intellectual development and is capable of high performance”, which means an intelligence quotient of at least two standard deviations above the mean on an individually administered standardized test of intelligence. For both of these outcomes, however, it is not enough to demonstrate disability or high intelligence, yet parents need to actively seek Individualized Education Plan (IEP) for their child. In that, both classifications are often the result of parent and teacher conferences that culminate in drafting such a plan and assigning child to appropriate disability/gifted group that we observe in our data.

For a limited set of cohorts who complete compulsory schooling in our data range, born in years 1992 and 1993, we also observe high school completion and their coursework in middle and high school. In this subset of observations, however, we cannot link siblings, and thus we are restricted to August vs. September comparisons of singleton births more generally. We define four high school graduation outcomes: graduating with a standard diploma, graduating with any diploma, not graduating on time but remaining in schooling, and not graduating on time and dropping out. The distinction between the two diploma types is that in the former case student graduates on time within four years, fulfilling all the requirements set out by Florida Department of Education, while in the latter group we include both standard diploma as well as GEDs, special diplomas for students with disabilities, and diplomas for other students who achieved a somewhat less rigorous set of coursework requirements. Therefore, the latter set includes diplomas with lower ability requirements. In addition to graduation outcomes, we also observe elective coursework for children in this sample. In middle school these are advanced and remedial courses in mathematics and reading, while in high school these include advanced placement (AP) courses. In the latter case, we can distinguish between following subjects: mathematics, English, science, social sciences and computer science.

We start with documenting demographic differences between the full population of births and the set of families whom we include in the empirical analysis (Table A1). First, it is worth noting that August and September births do not appear to differ substantially from all Florida births (columns 1 and 2) suggesting that seasonality in birth characteristics might be less of a problem in this analysis as compared to some other studies. That said, these averages may still mask important heterogeneities. Comparing columns 2 and 3 reveals the cost of only being able to utilize students attending public schools and remaining in these schools until at least third grade, where we first observe their test scores, as the sample used in the analysis is negatively selected compared to full population of births. Children observed in public schools are more likely to be African-American (25.8 percent vs. 22.4 percent), less likely to have college educated (15.2 percent vs. 20.1 percent) or married (60.7 percent vs. 65.2 percent) mother and more likely to utilize Medicaid payments during birth (50.8 percent vs. 45.1 percent). Most of these differences are due to the fact that more affluent families are more likely to send their children to private schools or leave the state than are

---

<sup>10</sup>Cognitive disabilities include: educable mentally handicapped, trainable mentally handicapped, language impaired, intellectual disability, profoundly mentally handicapped and developmentally delayed. Behavioral disabilities include emotionally handicapped, specific learning disabled, severely emotionally disturbed and autistic. Physical disabilities include orthopedically impaired, speech impaired, deaf or hard of hearing, visually impaired, hospital/home bound, dual sensory impaired, deaf and traumatic brain injury.

less affluent families, rather than any substantial additional selection occurring between school start and third grade.

More to the point of the present paper, it is also the case that fewer September-born children than August-born children are enrolled in public school at least through third grade. If the “missing” September children have particularly favorable or unfavorable academic achievement potential it could bias our school starting age estimates. The August-September gap in demographic characteristics among the full population and children included in the analysis is similar across most dimensions except for maternal education and poverty. On the other hand, even these differences are small and never exceed five percent of the mean value for a given characteristic.<sup>11</sup> That said, in Section 3 below we formally document these potential selection issues and carry out a bounding exercise to determine the degree to which they might influence our conclusions.

## 2.2 Methods

As mentioned above, it can be challenging to estimate the effects of school starting age, because a student’s age when entering primary education can be manipulated (via birth timing and/or redshirting) and may be correlated with family background characteristics. It has been shown that seasonal birth rates (which affects age relative to a cutoff) may vary based on family background characteristics (Buckles and Hungerman 2013). Research has also shown seasonal patterns in birth outcomes, mental health, neurological disorders, adult height, life expectancy, intelligence, and income (Currie and Schwandt 2013). There is evidence that conditions at conception, such as in utero exposure to illness/disease (Currie and Schwandt 2013) or nutrient deprivations due to seasonal nutritional intake (Barker 1990), may have an effect as well. Relatedly, we also know that parents can manipulate when children start school by redshirting. These redshirted children tend to be more likely male, white and from higher socioeconomic statuses (Bassok and Reardon 2013). As a consequence, comparing children based on their age when starting school is often fraught with omitted-variables concerns, and even results from studies with sufficient numbers of observations to make use of regression discontinuity evidence – say, comparing September births to August births in locales with a September 1st cutoff for school entry – may still be subject to omitted-variables bias due to endogenous birth timing.<sup>12</sup>

To address these challenges we proceed with the following empirical specifications. First, we begin with a simple model of the relationship between student outcomes and month of birth. In the main specification we restrict our attention to the August-September comparison, where September born children are about one year older than August born children at the time of school entry. For each child we only know year and month of birth, and thus we cannot perform more standard regression-discontinuity analysis with daily-level running variables. Therefore, we estimate the following equation:

$$Y_i = \beta Sept_i + \gamma X_i + \varepsilon_i \tag{1}$$

---

<sup>11</sup>In addition, comparing columns 3 and 6 in Table A1 demonstrates that the sample of siblings observed in Florida schools is modestly positively selected as compared to all students born in August or September and attending public schools. Children with siblings in our sample are more likely to have mothers who are college educated (19.9 percent vs. 15.2 percent) and married (63.4 percent vs. 60.7 percent) at the time of birth.

<sup>12</sup>Another challenge to estimate the effects of school starting age, summarized by Angrist and Pischke (2008) as a “fundamentally unidentified question” is that there is no way to decompose the effect of school starting age on an outcome measured during the schooling process into its three separate components: effect of a child’s age at school entry, effect of their age at the time of outcome measurement, and the effect of their age relative to their peer group. But it is also important to note that this deterministic link between the first two components disappears in a sample of adults past their schooling career as found in research such as Black et al. (2011).

where  $Y_i$  is one of the outcome variables for child  $i$  as defined in Section 2.1: kindergarten readiness; test scores in grades 3 to 8; being redshirted; being retained in an early grade; disability status; gifted status; middle and high school course selection; and high school graduation.  $Sept_i$  is an indicator variable for being born in the month of September;  $X_i$  contains mother and child control variables including year of birth dummies, maternal education, marital status at birth, medicaid paid birth, maternal race and ethnicity, child’s gender, log birth weight, gestational length, start of prenatal care in first trimester, and indicators for congenital anomalies, abnormal conditions at birth and maternal health problems;  $\varepsilon_i$  is the error term. In order to maintain as balanced sample as possible we estimate redshirting and retention behaviors for the population where we also observe test scores.<sup>13</sup>

In Equation 1 we do not include any demographic controls since we also present heterogeneity analyses utilizing these covariates. However, we do control for birth endowments of children as they may vary within a year (Currie and Schwandt 2013). The parameter of interest,  $\beta$ , is the causal effect of age under the assumption that the unobservables are not correlated with month of birth. The exogenous variation in school starting age comes from variation in month of birth (August vs. September) and the administrative school starting rule in Florida (September 1st), thus generating a fuzzy regression discontinuity design. The identifying assumption can be then translated into the following statement: children born in August and September are identical on observable and unobservable characteristics except for the age at which they begin schooling. In the case of Florida, akin to papers cited above, we also find that being born in September is correlated with observable family characteristics e.g. better educated and Hispanic mothers are less likely to have September births while mothers with Medicaid births are more likely to deliver in September. These differences are generally small – effect sizes between 0.2 percent for the African-American indicator and 3.8 percent for the college graduate mother indicator – but to further alleviate the endogeneity concerns we also propose a sibling fixed effects strategy.

In order to implement the fixed effects strategy, we first restrict the sample to families where we observe at least two siblings in our data. Then we further require that these siblings are first two in the family and both are born in either September or August. The estimating equation becomes:

$$Y_{ij} = \delta_j + \beta Sept_{ij} + \gamma X_{ij} + \varepsilon_{ij} \quad (2)$$

where  $Y$ ,  $Sept$ ,  $X$  and  $\varepsilon$  are defined as in Equation 1 but are now additionally subscripted with  $j$ , which indexes families. In Equation 2,  $\delta_j$  is a mother fixed effect that accounts for observable and unobservable characteristics that are shared by siblings and do not vary over time. Additional control in vector  $X$  is an indicator for being second born and the standard error  $\varepsilon$  is now clustered at the mother level for all outcomes. The identifying variation comes from the fact that one of the siblings is youngest and one is the oldest in their grades at school entry. Although an improvement over simple OLS, the potential endogeneity concerns that this strategy cannot resolve are any form of cross-sibling reinforcing/compensatory behavior or sibling spillovers (Black et al. 2017; Landerso et al. 2017a; Qureshi 2017). We directly investigate the former one by examining redshirting and retention. The latter is beyond the scope of this analysis; however, since we find remarkably similar academic achievement estimates across different samples and estimation strategies we suspect that this issue is an unlikely source of bias.

---

<sup>13</sup>We do not impose this restriction on kindergarten readiness because we do not have data for cohorts 1997 to 1999. The results are similar when we estimate the effects on redshirting, retention and test scores for all children for whom we can observe kindergarten readiness.

## 3 Results

### 3.1 Short- and medium-run outcomes

Table 1 documents the effect of school starting age on test scores, redshirting and retention for a variety of samples and two specifications. In each regression we compare September vs. August born children without (odd numbered columns) or with controls (even numbered columns). These additional covariates are described in Section 2.2. The main take home point of this table is that the point estimates are very similar regardless of the exact econometric specification used, which validates our regression discontinuity design. Furthermore, they are very similar for test scores but differ for the other two outcomes across different samples. In particular estimates for redshirting become more negative while for retention less negative as we move from sample of singletons (Panel A) to sibling sample (Panels B and C), and then further to siblings with the same parents (Panel D). These latter samples have higher SES which is evident not only when comparing mean test scores between Panels A and D but also by increasing redshirting and declining retention rates (Schanzenbach and Howard 2017). This finding and difference in estimates between test scores and remediation techniques as well as opposite movement of redshirting and retention preview our main heterogeneity result.

Returning to test score estimates, in Column 1 of Panel A we see that the September births score 0.197 SD higher than their August counterparts, and this estimate increases by only 0.005 when we add health and demographic controls.<sup>14</sup> In this analysis test scores are pooled across six grades and averaged for mathematics and reading, but in Figure 1 we show that estimates are about two times larger in grade 3 than they are in grade 8. However, even the latter at 0.158 SD is economically and statistically significant irrespective of exact econometric specification. Table A2 further documents that differences are modestly larger in reading than in mathematics. We next move to a specification in which we compare August and September births *within the same family*, by controlling for family fixed effects. We first confirm, in Panel B, that the OLS regression discontinuity estimates are essentially identical if we focus on the set of observed siblings relative to the full set of singletons; the point estimate is 0.216 SD for this sample, similar to the 0.197-0.202 SD estimated for the full population of singletons. When we actually control for family fixed effects in Panel C, we find the results are extremely similar – ranging from 0.216-0.218 SD – and when we choose an even more restrictive comparison, in which we estimate sibling fixed effects regression discontinuity models when both parents are the same for both siblings (Panel D), the estimates remain essentially unchanged, ranging from 0.222-0.223 SD.

In summary, while one might have been concerned that unobserved family characteristics for children born in September versus August might be driving observed differences in outcomes for September versus August births, the results from Table 1 make it clear that controlling for family characteristics and behavior does not substantially affect the estimated relationship between school starting age and test scores. We conclude ex post from this analysis that much of the regression discontinuity estimates in the literature are most likely not contaminated with quantitatively important family selection issues. The estimates for redshirting and retention are affected based on the estimation sample used. However, this difference is driven by substantial heterogeneity in the effects of being born in September vs. August for these outcomes across SES spectrum. For test scores, we do not detect such heterogeneity.

---

<sup>14</sup>Appendix Table A3 documents the OLS, reduced form and instrumental variable (using an indicator variable for September as an instrument for age at test) estimates for test scores based on the sample of singletons. The instrumental variables are not our preferred specification as the instrument likely does not satisfy the monotonicity assumption due to differential redshirting documented in Table 1 (Barua and Lang 2016). We provided them in the Appendix to give readers a sense of the difference in magnitudes between the IV and reduced form estimates.

In Section 2.1 we have noted that our sample consists only of children who attend public schools in Florida and stay in the system at least until third grade, the first time we observe test scores. Since this sample is positively selected and the selection correlates with being born in September (Table A4) the estimates presented in Table 1 may be biased.<sup>15</sup> To address this problem we propose a bounding exercise where we impute either 5th or 95th percentile of tests scores to students whom we either do not match to public schools or do not observe with test scores in public schools (for example because they leave the public schools between kindergarten and commencement of testing). These bounds are presented in square brackets in Table 1 and suggest that our preferred estimates are not substantively biased due to selection. The range of the bounds is also no greater than 6 percent of a standard deviation, that is about a fourth of the estimated effect in the most conservative approach.

In Figure 2, we examine the relationships found in Table 1 in more depth. In particular, we display the point estimates which come from a separate month-to-month comparisons using our larger sample of singletons on test scores, as well kindergarten readiness, early retention, and redshirting. We have not included kindergarten readiness estimates in Table 1 because due to data limitations these cannot be estimated in siblings sample. In Panel A we observe that, regardless of which month-to-month comparison we employ, the older children of the pair are more likely to be ready for kindergarten at the start of formal schooling. However, in all cases except for the September versus August comparison, the estimated differences are small, albeit often significantly distinct from zero. On the other hand, in the case of the September versus August discontinuity, the difference is dramatically larger than seen elsewhere – an older-child advantage of 10 percentage points – over five times higher than the second-largest difference. For test scores, reported in Panel B, the September versus August estimate is 0.17 SD larger than second-largest difference (0.20 SD vs 0.03 SD).

Panels C and D of Figure 2 show the differential effects of being older on the probability of being redshirted (Panel C) or being retained in early grades (Panel D). Here we find that the September versus August difference in redshirting rates (5 percentage points) is more than double the next largest month to month comparison. Parents redshirt children born in both July and August but roughly twice as many August babies are redshirted than those born in July. Regarding early-grade retention (Panel D), the point estimate for the September versus August comparison is -0.152 and dwarfs any of the month-by-month comparisons. Therefore, Figure 2 gives us much confidence that our fuzzy regression discontinuity design is accurately picking up the important age differences in our data.<sup>16</sup>

We next move to other educational and health outcome measures. In Table 2 we explore the effect of school starting age on disability and gifted status. Columns (1) and (2) show effects on any type of disability, and we find that September births have 4.6 percentage points lower probability of having disability label than their August counterparts. This result is confirmed in sibling fixed effects analysis and is invariant to including additional controls. Decomposing the effect by disability type (columns (3) to (8)) we show that in singletons sample the estimates are largest for behavioral and physical disability while in sibling fixed effects analysis these are only statistically significant for the former group.<sup>17</sup>

---

<sup>15</sup>We formally document this selection in Table A4, where the dependent variables are either being matched between birth and public school records or being observed with third grade test scores conditional on being merged to public school records. Since the sibling match occurred via school records, this particular analysis can only be done for the latter selection. Regardless of the specification, we find that September born children are about 2 percentage points less likely to be merged between birth and school records and are between 0.3 and 0.6 percentage points more likely to be included in the empirical sample conditional on being merged between the two data sources.

<sup>16</sup>Sibling fixed effects results for panels B to D are qualitatively very similar but have larger standard errors due to decreased sample sizes. This is consistent with findings reported in Table 1

<sup>17</sup> Sample sizes vary by disability type because we always compare children with a given disability type to healthy

The exact mechanism behind the age effect in disability is unclear to us. On the one hand, it may be due to mislabeling cognitive and non-cognitive immaturity among young for grade children as disability symptoms. These children are biologically younger at school entry, but they are held to the same academic standards as their older counterparts. Thus, we might expect differential classification rates by age if educators and parents pursue a disability assessment for their children who academically achieve at lower levels. On the other hand, we cannot rule out a direct effect of being young for grade on disabilities, especially behavioral ones, where a child could struggle due to peer pressure and relative ranking among their classmates. Irrespective of the exact cause our estimates are of policy relevant magnitude e.g. result in column (2) of Panel A implies effect size of 19 percent. They are also concordant to the literature on ADHD over-diagnoses in young at school entry children (Elder 2010; Evans et al. 2010; Morrow et al. 2012) but bolster these findings with within-family design and health-at-birth controls, both of which could be important econometrically. Finally, in columns 9 and 10 we further explore a potentially positively perceived IEP outcome - gifted status. These results suggest that old for grade students are more likely to be labeled as gifted, which again could be either due to superior intellectual development or the desire of parents to label their over-performing children.

For cohorts born in 1992 and 1993, where we cannot implement the sibling fixed effects design but where the children are old enough to conclude compulsory schooling, we can observe additional outcomes. Table 3 explores these medium-run outcomes that to our knowledge have not been explored in the literature thus far. We estimate the August-September difference in taking advanced or remedial courses in middle school or Advanced Placement courses in high school. Advanced courses such as ones offered in the Advanced Placement Program were designed to provide high school students a way to learn university level material while in high school and serve as an important signal in college admissions (Klopfenstein and Thomas 2009). Furthermore, there are studies showing that passing AP exam scores are strong predictors of success at university (Hargrove et al. 2008; Keng and Dodd 2008).

In Table 3 we observe a large August-September difference in these non-test score outcomes. In particular, we find negative effects for remedial courses in middle school. Conversely, we find positive effects of having September birth on middle school advanced courses and all AP courses except computer science, which has very few students taking this class overall. Adding a large variety of demographic and health controls in Panel B makes little difference in terms of magnitudes and significance. These large differences may be surprising given that some of the previous literature has suggested that the age effects dissipate quickly and are not economically significant in later years (e.g. Elder and Lubotsky 2009).

Our final medium-run outcomes relate to high school graduation. We coded four variables in this domain: graduated and received a standard diploma, graduated and received any diploma (including a GED degree), not graduated but still in school more than five years after starting grade nine, and not graduated but has dropped out of school. In Table 4, odd numbered columns do not include any controls while even numbered columns control for health and demographic covariates. The August-September difference for graduating and receiving a standard diploma is positive and statistically significant regardless of specification, however, we do not find any other consistent results across the additional outcomes. Overall, our high school graduate findings are inconsistent with findings from Dobkin and Ferreira (2010), Cook and Kang (2016), Hemelt and Rosen (2016) and Tan (2017). This can be potentially explained by two opposing forces in action at the same time when measuring the August-September difference – both that the September-born students have a cognitive advantage over their August counterparts (as can be seen proxied by their test scores) and also that they

---

children where both groups are born either in August or September.



have the ability to dropout of high school for a longer period of time due to their increased age. It appears that in our sample at least for the most positive outcome, unlike in some previous research, the September-born children’s increased cognition is the dominating force.

### 3.2 Heterogeneity

A majority of the previous research has offered few and conflicting insights in terms of heterogeneity in the August-September differences. For example, some papers find larger differences for girls (Datar 2006) while other for boys (Puhani and Weber 2007; McEwan and Shapiro 2008). Similarly there is evidence that effects are larger among higher SES families in some contexts (Elder and Lubotsky 2009; Tan 2017) but in lower SES families in others (Datar 2006; Black et al. 2011; Cook and Kang 2016; Hemelt and Rosen 2016). Because of the contradictory results in the literature examining effect heterogeneity, especially using large-scale linked administrative data, is important as it may provide further insights on these conflicting previous results. We have already hinted in Section 3.1 that heterogeneous effects may further depend on the outcome under scrutiny. The Florida data are particularly suited to explore heterogeneity in great detail, as these include an incredibly detailed information on a highly diverse population with over 20 percent of African-American, Hispanic and high school dropout families. In the analysis that follows, we investigate the degree to which estimated effects of school starting age vary by race/ethnicity, maternal education, family poverty, birth weight, gestational age, school quality, and sex. In particular, the interaction between initial endowments and school starting age has never been studied before to our best knowledge, and appears crucial from the policy perspective given the hypothesized interaction between early childhood inputs (Cunha et al. 2010).

We present the heterogeneity results in Figures 3-10. In each figure, the bar or dot represents a point estimate and it includes a 95 percent confidence interval (whiskers) from our September versus August singletons regression discontinuity comparison.<sup>18</sup> As seen in Figure 3, Panel A, the September-August difference in kindergarten readiness is much lower for high-SES families than for low-SES families (whether measured by family income proxied by Medicaid payment or maternal education groups); and much lower for white families than for minority families.<sup>19</sup> These are exactly the groups that also experience higher redshirting rates. On the other hand, differences in readiness are comparatively low for higher-birth weight infants relative to lower-birth weight infants or full-term infants relative to premature or post-term infants suggesting no interaction between initial health endowments and age at the start of education (see Figures 4 and 5).

Remarkably, as seen in Figures 3-6, the estimated effects of school starting age on test scores are highly similar across a wide range of SES groups as well as a wide range of initial infant health, or a wide range of school quality.<sup>20</sup> These findings indicate that school starting age affects children’s test scores by essentially the same amount – despite the fact that different groups of families have children with different average health at birth or academic achievement and are differentially proactive regarding how they attempt to remediate their young-for-grade children.

Differences in early family remediation behaviors can help to explain why we document considerable heterogeneity in kindergarten readiness but not in third grade test scores by different family

---

<sup>18</sup>Our sibling fixed effects heterogeneity results are again qualitatively similar; however, due to small sample sizes we often lose statistical power. In order to facilitate comparability in heterogeneity estimates we drop all controls in these analyses but as documented in Section 3.1 they do not matter for our average estimates.

<sup>19</sup>Elder and Lubotsky (2009) also find significant heterogeneity during the fall of kindergarten but they find larger age effects for the children from higher socioeconomic status families, which is at odds with our estimates.

<sup>20</sup>We are unable to explore differences in kindergarten readiness or redshirting practices stratified by school quality as these two outcomes are measured at the very beginning of schooling, and thus cannot be affected by the quality of school that a child attends in the first grade.

background groups. In that, we postulate that remediation behaviors might be partially responsible for the presence of heterogeneity at the start of school but not in subsequent test scores as we observe that high-SES families are more likely to redshirt their August-born children, while children from low-SES families are more likely to be retained in early grades. Importantly, while redshirting has the potential of affecting both kindergarten readiness and subsequent test scores by the nature of school retention it happens only after a child starts schooling, and thus cannot have an effect on kindergarten screening results. This difference in timing is consistent with the pattern observed in the data, and the two approaches to remediating young for grade children may be the cause of the sharply reducing SES-age profile for August-born versus September-born children by third grade. Later in this paper we provide some suggestive evidence regarding the potential efficacy of these remediation strategies.

Exploring heterogeneity further, we look to the student’s sex. Boys are redshirted more often than girls (Bassok and Reardon 2013; Schanzenbach and Howard 2017), implying that many families think that school starting age is more relevant for their sons than for their daughters. In Figure 3, we graph the point estimates for males and females in our sample. In terms of kindergarten readiness, we find that September males have a larger age advantage than September females as compared to August births. However, in terms of averaged test scores, we are unable to statistically distinguish between male and female estimates – they are equally as big, around 0.2 SD. At the same time, there are significant gender differences in behaviors of parents and schools in terms of redshirting and retention. Male August babies are significantly more likely to be redshirted than female August babies, perhaps due to “conventional wisdom” regarding gender differences in maturity, or perhaps due to the fact that August-born boys are somewhat less ready to start school than are August-born girls. August-born boys are also differentially more likely to be retained in early grades (relative to their September-born counterparts) than are August-born girls.

We further examine the stratification by socioeconomic status and gender and provide each heterogeneity estimate separately for boys and girls (see Autor et al. (2016a) and Autor et al. (2016b) for an in depth exploration of gender-SES gaps in Florida). These results can be found in Table A5. In Panel A, we find that across all categories, the kindergarten readiness gap between September versus August-born children is larger for males than females. When examining the average test score gap in Panel B, we find that the test score gap is similar between males and females except for the children with college educated mothers and mothers who were not on Medicaid. In these cases, the test score gap is actually larger for females. We find that the August-born males are redshirted and retained more in all categories but the magnitude of redshirting is substantially higher for the boys with mothers who are college graduates, non-Medicaid, or white. These facts together indicate that the increased prevalence of redshirting might help to boost test scores as we have seen that these males are also the children that have a smaller September-August test score gap.

We next move on to our other medium-run outcome measures: disability and gifted status (Figure 7); middle school course enrollment (Figure 8); high school course enrollment (Figure 9); and high school graduation outcomes (Figure 10). In each of these figures, we consider three cuts of the data: by education levels of the mother; by race/ethnicity; and by gender. We are also able to investigate differences by income for disability and gifted status but not for the other outcomes as the income measure is not available for those particular cohorts. We do not find much of a heterogeneity across birth weights, gestational age, and school quality, and thus for brevity we do not report these results.

In Figure 7 we find a striking education gradient and corresponding income gradient where higher SES families seem to be able to mitigate some of the school entry age effect on disability identification, and these are especially pronounced for behavioral problems which may be particularly affected by relative age effects. Furthermore, we find evidence for gender differences, with males being more elastic, which again is particularly pronounced in the behavioral domain. These heterogeneity results

are similar to what has been documented for ADHD by [Evans et al. \(2010\)](#) and [Elder \(2010\)](#), and may be due to differences across these groups in parent’s demand for disability assessment and identification or by differential access to medical care and school psychological resources ([Currie and Gruber 1996](#)). At the same time, we find no statistically significant differences in terms of gifted status by education, income, race/ethnicity, or gender in [Figure 7](#), and we think that if “labeling desire” would be a dominating effect, we should then also observe a gradient in this potentially positive IEP measure.

Turning to course enrollment, we find SES gradients in both middle school ([Figure 8](#)) and in high school ([Figure 9](#)). Although, we do not find much of a gender gap in middle-school there is a gap between boys and girls in all AP courses except for math and computer science, with females having larger August-September differences in AP enrollment than males. On the other hand, differences based on socioeconomic variables are generally more pronounced in middle school rather than in high school. Finally, we do not find statistically significant heterogeneous effects of the school starting age on our high school graduation outcomes ([Figure 10](#)), however, these are relatively imprecise and in [Panel A](#), for instance, the difference among children of college educated mothers is visibly smaller as compared to other education groups.

Summarizing our heterogeneity analysis, we observe that there exists very little heterogeneity in the August-September difference on test scores across a substantial array of different child, family, and school dimensions, despite pronounced age effects in kindergarten readiness. We do, however, find heterogeneity in other non-test score outcome measures such as disability identification and course selection in middle and high school. It seems that the outcomes that display heterogeneity are measures that can be influenced the most by parental involvement or intervention, but this relationship is speculative at best. Moreover, these findings do not provide evidence regarding which remediation mechanism, if any, leads to heterogeneous effects of school starting age – just that heterogeneity in estimated effects exists for some outcomes and not for others. Importantly though, we are able to investigate a broad range of outcomes including some that have never been studied before. In the following sections of this paper we attempt to uncover whether school policies or remediation efforts could be responsible for some of these patterns of results.

### 3.3 Interaction between school policies and school starting age

One of the more challenging aspects of the school entry literature is that policy recommendations are generally hard to come by despite the stark differences in outcomes of children who enter school early versus late. It is difficult to imagine the administration of a school system in which there were no school entry cutoff dates that causes the age distribution of children at the entry. It is possible to decrease these age differences, however, by having a more staggered school entry such that new primary aged students enter school either in the fall or in the spring depending on birth dates. This requires multiple classrooms of the same grade in each school, which is not feasible in many locales, and which might make other class-composition policies harder to execute. In addition, there exists much speculation on the exact mechanisms at work behind the measured age differences. These include the age effects being entirely driven by differences in skill accumulation prior to kindergarten ([Elder and Lubotsky 2009](#)), or driven by differences in actual biological age at test time outweighing the position in the age distribution ([Black et al. 2011](#); [Cascio and Schanzenbach 2016](#)), or driven by differences in educational trajectories due to individuals in authority positions such as teachers/coaches mistaking maturity with ability ([Bedard and Dhuey 2006](#)), or driven by these individuals in authority positions relating their evaluations of a child’s development to the child’s location in the age distribution ([Elder 2010](#)).

Because changing the administrative need for school entry cutoff dates seems unlikely, we explore

other common school policies to understand if there are any interactions between these policies and the magnitudes of the estimated effects of school starting age. In Table 5, we examine twenty different school policies that we observe occurring in Florida primary schools to understand whether any of these policies either mitigate or intensify the estimated September-August difference. We utilize the information on school policies and practices by interacting the presence of a given policy/practice with an indicator for September birth and further controlling for both dummy variables. This interaction term describes whether August vs. September gaps in test scores are larger or smaller in schools that have and do not have a given policy in place. Since our first achievement measure is based on performance in grade three we use third grade test scores as outcomes and policies that a child experiences in grade one as treatment.<sup>21</sup> Each row includes estimates using a different school policy. Column (1) analyzes each policy one at a time; whereas column (2) jointly includes all policy interactions and indicators. Panel A focuses on policies relating to before and after care. Panels B and C relate to schedules and staffing and to extending overall instructional time, respectively. Panel D measures class size whereas Panel E includes policies in place to improve the achievement of low-performing students.

What is notable is that only three policies consistently influence the estimate of the September-August difference: block scheduling, summer school requirement for grade advancement, and class size. Block scheduling refers to the practice when pupils have fewer but longer classes each day. Summer school for grade advancement is an additional coursework requirement for low-performing students to advance to next grade. It is worth noting that both policies are fairly common in use, at 36.5 percent and 58.9 percent respectively, and both exacerbate rather than ameliorate the August-September gaps. Furthermore, we also find that increasing class size is associated with larger achievement differences between old and young for grade students. This makes sense to us because larger classes are more likely to be heterogeneous, and thus putting young for grade children at relatively bigger distance in ability or maturity as compared to their peers. Overall, we view these estimates as not suggestive of any particular method that could alleviate the age-for-grade disparities among children. In Section 3.4, we move from strictly school level policies to remediation practices that can be partially impacted by parental decisions.

### 3.4 Exploratory analysis: Potential consequences of redshirting and early grade retention

Providing causal evidence on the effect of redshirting is difficult, as children who are being redshirted undoubtedly come from families that are different on both observables and unobservables. For example, in our sample, families with college-educated mothers are more likely to redshirt their children in comparison to families where the mothers did not complete college education. Thus, it is challenging to disentangle the act of redshirting from the observable and unobservable qualities of these families. Based on the heterogeneity studied in Section 3.2, we concluded that redshirting might potentially increase the test scores of those being redshirted - primarily males from higher socioeconomic status families. Below we provide additional associational evidence on this relationship.

Schools and school districts have considerable leeway on the rules and regulations set in their district regarding who is allowed to be redshirted. Some districts allow for large levels of redshirting whereas others do not. Also, different “parenting cultures” and even daycare prices may affect redshirting practices which in turn may affect redshirting levels across school districts and within school districts over time. As a consequence, across the 65 (out of 67 total) county-level school

---

<sup>21</sup>As explained in Section 2.1 we use the initial survey responses from school year 1999-2000 as a permanent feature of the school and assign it to all their first graders over time. To the extent that school policies change over time our estimates are more noisy than if we were to observe policy variable every school year.

districts in Florida where we could construct this measure, redshirting rates vary from 0 to 8.5 percent, and August-birth redshirting rates range from 0 to 50 percent.<sup>22</sup> School districts vary dramatically in terms of early-grade retention rates as well. Early-grade retention rates range from 9.1 to 48.3 percent across the 65 school districts, and early-grade retention rates for August births range from 0 to 100 percent.<sup>23</sup> This variation is not due to observed background differences: if we were to predict redshirting and early-grade retention using the variables observed on a child’s birth certificate, we would have only expected to see a range of 1.0 to 2.6 in county-level differences in redshirting rates and a range of 14.4 to 26.7 in county-level differences in early-grade retention rates.<sup>24</sup> Finally, districts that pursue one policy or practice do not necessarily pursue the other policy or practice. For example, district-cohort correlation between redshirting and retention rates is 0.657.

In Table 6, we examine the relationships between school district-level differences in the rates of redshirting as well as early-grade retention and test scores. We collapse the data at year of birth×school district×September birth level. In the first column of Panel A, we regress average test scores on the fraction of children redshirted, an indicator for a September birth and the interaction of these two variables. We also include school district and cohort fixed effects and cluster our standard errors at the school district level. We find that the percent redshirted is positively related to the average test score level but that the interaction between percent redshirted and being born in September is negatively related. This indicates that the school districts that have higher proportions of redshirted children have lower September/August test score gaps. In the first column of Panel B, we regress average test scores on the fraction of children retained by the school in early grades. Here we find that retention is negatively related to test scores and that the interaction is also negatively related. This implies that school districts that have higher levels of retention in early grades also have lower old versus young test score gaps. This evidence is necessarily only suggestive because despite the school district and cohort fixed effects there may still be unobservable variables that affect both the level of redshirting/retention and the level of test scores. Nonetheless, this district level evidence paired with previous individual level analyses gives us confidence to lean in the direction of saying that school districts where redshirting and early grade retention are more prevalent also have smaller September-August gaps in test scores.

We further investigate this question by considering heterogeneity in *whose* test scores are differentially related to school district-level redshirting and early grade retention rates.<sup>25</sup> The remaining columns of Table 6 help to tell this story. While only correlational in nature, we find that in school districts where one remediation strategy is especially prevalent, September-August perfor-

---

<sup>22</sup>We exclude two counties from this analysis because we do not observe children’s place of residence at birth for those born in 1994 and 1995 in these counties. These two counties constitute 1.5 percent of the full population of births in years 1996 to 2000. Our results are fundamentally unchanged when we use all 67 school districts and limit birth cohorts to 1996 to 2000, when we observe location of birth for the entire state.

<sup>23</sup>Some of the school districts in our sample are very small and have less than 10 students born in a given year and month. If we restrict our sample to counties with at least 50 August births in each year we are left with 29 school districts, and then the August redshirting rates range from 1.3 to 18.1 while August early retention rates range from 12.1 to 44.3.

<sup>24</sup>We regress at individual level indicator for being redshirted or early retained on infant gender, month and year of birth dummies, birth weight, gestational age, dummies for congenital anomalies and abnormal conditions at birth as well as on mom’s race, ethnicity, education, foreign born status, medicaid paid birth, health problems and start of prenatal care in first trimester. Then, we use coefficients from this regression to predict values of the dependent variables and collapse them at school district (65 districts) and year level (7 years).

<sup>25</sup>Here, we limit our analysis to school districts with at least five children in each cell (year of birth by September birth heterogeneity groups). This restriction yields unbalance repeated-cross section of observations. Retention results are similar when we impose full panel restriction across heterogeneity dimensions and years while the results for redshirting become less precisely estimated.

mance gaps tend to be lower for the demographic and socioeconomic groups that in general are less likely to experience that type of remediation. In the case of redshirting, the largest reductions in the September-August performance gaps associated with district-level redshirting rates are for children with high school dropout mothers and for children who were born in poverty, as indicated by Medicaid-funded births. In the case of early grade retention, the largest reductions in the September-August performance gaps associated with district-level early grade retention rates are for children with high school graduate and college graduate mothers; for those whose births were *not* funded by Medicaid; and for white students rather than minority students. This pattern of findings provides further evidence that while redshirting and early grade retention are remediation tools that could have negative consequences, such as those described by [Schanzenbach and Howard \(2017\)](#), there are potential vehicles for remediation inside and outside of the school system – especially for groups for whom the remediation strategy is less frequently used. However, more experimentation and causal evidence is necessary before we are prepared to make this recommendation.

## 4 Conclusions

In this paper, we document, using matched administrative data from the state of Florida, the most robust to date evidence on the short- and medium-run effects of school starting age on children’s cognitive development. The regression discontinuity approach as well as the month-to-month within family sibling fixed effects comparison where we control for all the time invariant endowments and family characteristics show that September born children benefit developmentally in comparison to August born children. Our test score findings are very similar irrespective of the empirical approach chosen, which suggests that much of the regression discontinuity estimates in the literature thus far are most likely not contaminated with quantitatively important family selection issues.

We find heterogeneity in terms of kindergarten readiness along with disability status and middle- and high school course selection. But we also document a striking lack of heterogeneity in test scores and high school graduation rates by student, maternal, and school characteristics. At the same time, we observe different compensatory behaviors targeted towards children from different socioeconomic statuses who are youngest in their schooling cohort. While the more affluent families tend to redshirt their children to give them competitive advantage, families that are unable to do this - either due to lack of awareness or resources - are surrogated by the schooling system, which retains their children in grades prior to testing. This differential remediation also helps explaining why we find larger kindergarten readiness gaps for lower SES children that then vanish at the time of testing. Namely, since low SES children are not redshirted but rather retained there is no scope for retention to affect children’s cognitive development prior to the start of schooling. Together, both of these mechanism seem to be equally effective because children coming from different socioeconomic backgrounds end up at roughly the same educational levels at the time of testing irrespective of the affluence.

We have also explored if particular school policies can ameliorate the September-August cognitive gaps. We find that the practices of block scheduling and summer school requirements for grade advancement among low-performing students are associated with larger rather than smaller school entry age effects. Therefore, these policies should be carefully considered by schools if the goal is to decrease the magnitude of age effects for their students. We did not find differential influence of any other policies but smaller classrooms in first grade appear to shrink the achievement gap between youngest and oldest children in the classroom. Finally, we also explored whether the relationship between remediation techniques and test scores estimated at individual level translated into policy relevant district-level variation. We show that the percent of children redshirted is positively related to the average test score level but that the interaction between the percent redshirted and being

the oldest in a cohort is negatively related. At the same time, retention is negatively related to test scores but the interaction between retention and being the oldest in a cohort is also negatively related. Together, these findings indicate that school districts where redshirting and early grade retention are higher have smaller relative age gaps in test scores.

## References

- Angrist, Joshua D and Jörn-Steffen Pischke**, *Mostly harmless econometrics: An empiricist's companion*, Princeton university press, 2008.
- Attar, Itay and Danny Cohen-Zada**, “The effect of school entrance age on educational outcomes: Evidence using multiple cutoff dates and exact date of birth,” IZA Discussion Paper 10568, 2017.
- Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman**, “Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes,” NBER Working Paper 22267, 2016.
- , –, –, –, and –, “School Quality and the Gender Gap in Educational Attainment,” *American Economic Review Papers and Proceedings*, 2016, 106 (5), 289–295.
- Barker, David**, “The fetal and infant origins of adult disease,” *BMJ*, 1990, 301 (6761), 1111.
- Barua, Rashmi and Kevin Lang**, “School entry, educational attainment and quarter of birth: A cautionary tale of a local average treatment effect,” 2016.
- Bassok, Daphna and Sean Reardon**, “Academic redshirting in kindergarten: Prevalence, patterns, and implications,” *Educational Evaluation and Policy Analysis*, 2013, 35 (3), 283–297.
- Bedard, Kelly and Elizabeth Dhuey**, “The persistence of early childhood maturity: International evidence of long-run age effects,” *Quarterly Journal of Economics*, 2006, 121 (4), 1437–1472.
- Bharadwaj, Prashant, Petter Lundborg, and Dan-Olof Rooth**, “Birth weight in the long run,” *Journal of Human Resources*, 2017, *forthcoming*.
- Black, Sandra, Paul Devereux, and Kjell Salvanes**, “From the cradle to the labor market? The effect of birth weight on adult outcomes,” *Quarterly Journal of Economics*, 2007, 122 (1), 409–439.
- , –, and –, “Too young to leave the nest? The effects of school starting age,” *Review of Economics and Statistics*, 2011, 93 (2), 455–467.
- , **Sanni Breining, David Figlio, Jonathan Guryan, Helena Nielsen Skyt, Jeffrey Roth, and Marianne Simonsen**, “Sibling spillovers,” NBER Working Paper 23062, 2017.
- Buckles, Kasey and Daniel Hungerman**, “Season of birth and later outcomes: Old questions, new answers,” *Review of Economics and Statistics*, 2013, 95 (3), 711–724.
- Cascio, Elizabeth and Diane Whitmore Schanzenbach**, “First in the class? Age and the education production function,” *Education Finance and Policy*, 2016, 11 (3), 225–250.
- Cook, Philip and Songman Kang**, “Birthdays, schooling, and crime: Regression-discontinuity analysis of school performance, delinquency, dropout, and crime initiation,” *American Economic Journal: Applied Economics*, 2016, 8 (1), 33–57.
- and –, “The school-entry-age rule affects redshirting patterns and resulting disparities in achievement,” NBER Working Paper 24492, 2018.
- Crawford, Claire, Lorraine Dearden, and Costas Meghir**, “When you are born matters: The impact of date of birth on child cognitive outcomes in England,” London: Centre for the Economics of Education, 2007.
- , –, and –, “When you are born matters: the impact of date of birth on educational outcomes in England,” May 2010.
- Cunha, Flavio, James Heckman, and Susanne Schennach**, “Estimating the technology of cognitive and noncognitive skill formation,” *Econometrica*, 2010, 78 (3), 883–931.
- Currie, Janet and Hannes Schwandt**, “Within-mother analysis of seasonal patterns in health



- at birth,” *Proceedings of the National Academy of Sciences*, 2013, *110* (30), 12265–12270.
- **and Jonathan Gruber**, “Health insurance eligibility, utilization of medical care, and child health,” *The Quarterly Journal of Economics*, 1996, *111* (2), 431–466.
- Datar, David**, “Does delaying kindergarten entrance give children a head start?,” *Economics of Education Review*, 2006, *25* (1), 43–62.
- Dee, Thomas and Hans Sievertsen**, “The gift of time? School starting age and mental health,” *Health Economics*, 2017, *forthcoming*.
- Deming, David and Susan Dynarski**, “The Lengthening of Childhood,” *Journal of Economic Perspectives*, 2008, *22* (3), 71–92.
- Depew, Briggs and Ozkan Eren**, “Born on the wrong day? School entry age and juvenile crime,” *Journal of Urban Economics*, 2016, *96*, 73–90.
- Dhuey, Elizabeth and Stephen Lipscomb**, “What makes a leader? Relative age and high school leadership,” *Economics of Education Review*, 2008, *27* (2), 173–183.
- **and –**, “Disabled or young? Relative age and special education diagnoses in schools,” *Economics of Education Review*, 2010, *29* (5), 857–872.
- Dobkin, Carlos and Fernando Ferreira**, “Do school entry laws affect educational attainment and labor market outcomes?,” *Economics of Education Review*, 2010, *29* (1), 40–54.
- Du, Qianqian, Huasheng Gao, and Maurice D Levi**, “The relative-age effect and career success: Evidence from corporate CEOs,” *Economics Letters*, 2012, *117* (3), 660–662.
- Elder, Todd and Darren Lubotsky**, “Kindergarten entrance age and children’s achievement: Impacts of state policies, family background, and peers,” *Journal of Human Resources*, 2009, *44* (3), 641–683.
- Elder, Todd E**, “The importance of relative standards in ADHD diagnoses: evidence based on exact birth dates,” *Journal of health economics*, 2010, *29* (5), 641–656.
- Evans, William, Melinda Morrill, and Stephen Parente**, “Measuring inappropriate medical diagnosis and treatment in survey data: The case of ADHD among school-age children,” *Journal of Health Economics*, 2010, *29* (5), 657–673.
- Fertig, Michael and Jochen Kluge**, “The effect of age at school entry on educational attainment in Germany,” IZA Discussion Paper 1507, 2005.
- Figlio, David, Jonathan Guryan, Krzysztof Karbownik, and Jeffrey Roth**, “The effects of poor neonatal health on children’s cognitive development,” NBER Working Paper 18846, 2013.
- , – , – , **and –**, “The effects of poor neonatal health on children’s cognitive development,” *American Economic Review*, 2014, *104* (12), 3921–3955.
- , – , – , **and –**, “Long-term cognitive and health outcomes of school-aged children who were born late-term vs. full-term,” *JAMA Pediatrics*, 2016, *170* (8), 758–764.
- Fredriksson, Peter and Bjorn Ockert**, “Life-cycle effects of age at school start,” *Economic Journal*, 2014, *124* (579), 977–1004.
- Garfield, Craig, Krzysztof Karbownik, Karna Murthy, Gustave Falciglia, Jonathan Guryan, David Figlio, and Jeffrey Roth**, “Educational Performance of Children Born Prematurely,” *JAMA Pediatrics*, 2017, *171* (8), 1–7.
- Hargrove, Linda, Donn Godin, and Barbara Dodd**, “College Outcomes Comparisons by AP and Non-AP High School Experiences. Research Report No. 2008-3.,” *College Board*, 2008.
- Hemelt, Steven and Rachel Rosen**, “School entry, compulsory schooling, and human capital

- accumulation: evidence from Michigan,” *B.E. Journal of Economic Analysis and Policy*, 2016, 16 (4), 1–29.
- Hurwitz, Michael, Jonathan Smith, and Jessica Howell**, “Student age and the collegiate pathway,” *Journal of Policy Analysis and Management*, 2015, 34 (1), 59–84.
- Kawaguchi, Daiji**, “Actual age at school entry, educational outcomes, and earnings,” *Journal of the Japanese and International Economies*, 2011, 25 (2), 64–80.
- Keng, Leslie and Barbara G Dodd**, “A comparison of college performances of AP and non-AP student groups in 10 subject areas,” 2008.
- Klopfenstein, Kristin and M Kathleen Thomas**, “The link between advanced placement experience and early college success,” *Southern Economic Journal*, 2009, pp. 873–891.
- Landerso, Rasmus, Helena Skyt Nielsen, and Marianne Simonsen**, “How going to school affects the family,” Department of Economics Aarhus University, 2017.
- , – , and – , “School starting age and the crime-age profile,” *Economic Journal*, 2017, *forthcoming*.
- Larsen, Erling Røed and Ingeborg F Solli**, “Born to run behind? Persisting birth month effects on earnings,” *Labour Economics*, 2017, 46, 200–210.
- Lubotsky, Darren and Robert Kaestner**, “Do skills beget skills? Evidence on the effect of kindergarten entrance age on the evolution of cognitive and non-cognitive skill gaps in childhood,” *Economics of Education Review*, 2016, 53, 194–206.
- McAdams, John**, “The effect of school starting age policy on crime: Evidence from U.S. micro-data,” *Economics of Education Review*, 2016, 54, 227–241.
- McCrary, Justin and Heather Royer**, “The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth,” *American Economic Review*, February 2011, 101 (1), 158–95.
- McEwan, Patrick and Joseph Shapiro**, “The benefits of delayed primary school enrollment. Discontinuity estimates using exact birth dates,” *Journal of Human Resources*, 2008, 43 (1), 1–29.
- Morrow, Richard, Jane Garland, James Wright, Malcolm Maclure, Suzanne Taylor, and Colin Dormuth**, “Influence of relative age on diagnosis and treatment of attention-deficit/hyperactivity disorder in children,” *CMAJ*, 2012, 184 (7), 755–762.
- Muhlenweg, Andrea M. and Patrick A. Puhani**, “The evolution of the school-entry age effect in school tracking system,” *Journal of Human Resources*, 2010, 45 (2), 407–438.
- Muller, Daniel and Lionel Page**, “Born leaders: Political selection and the relative age effect in the US Congress,” *Journal of the Royal Statistical Society: Series A*, 2016, 179 (3), 809–829.
- Nam, Kigon**, “Until when does the effect of age on academic achievement persist? Evidence from Korean data,” *Economics of Education Review*, 2014, 40, 106–122.
- Ozek, Umut**, “Hold back to move forward? Early grade retention and student misbehavior,” *Education Finance and Policy*, 2015, 10 (3), 350–377.
- Pena, Pablo**, “Creating winners and losers: date of birth, relative age in school, and outcomes in childhood and adulthood,” *Economics of Education Review*, 2017, 56, 152–176.
- Puhani, Patrick and Andrea Weber**, “Does the early birth catch the worm? Instrumental variables estimates of early educational effects of age of school entry in Germany,” *Empirical Economics*, 2007, 32 (2), 359–386.
- Qureshi, Javaeria**, “Siblings, teachers and spillovers in academic achievement,” *Journal of Human*

*Resources*, 2017, *forthcoming*.

**Robertson, Erin**, “The effects of quarter of birth on academic outcomes at the elementary school level,” *Economic of Education Review*, 2011, *30* (2), 300–311.

**Rouse, Cecilia, Jane Hannaway, Dan Goldhaber, and David Figlio**, “Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure,” *American Economic Journal: Economic Policy*, 2013, *5* (2), 251–281.

**Schanzenbach, Diane Whitmore and Stephanie Larson Howard**, “Season of birth and later outcomes: Old questions, new answers,” *Education Next*, 2017, *17* (3), 18–24.

**Schneeweis, Nicole and Martina Zweimuller**, “Early tracking and the misfortune of being young,” *Scandinavian Journal of Economics*, 2014, *116* (2), 394–428.

**Schwerdt, Guido, Martin West, and Marcus Winters**, “The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida,” NBER Working Paper 21509, 2015.

**Smith, Justin**, “Can regression discontinuity help answer an age-old question in education? The effect of age on elementary and secondary school achievement,” *The B.E. Journal of Economic Analysis and Policy*, 2009, *9* (1), 1–30.

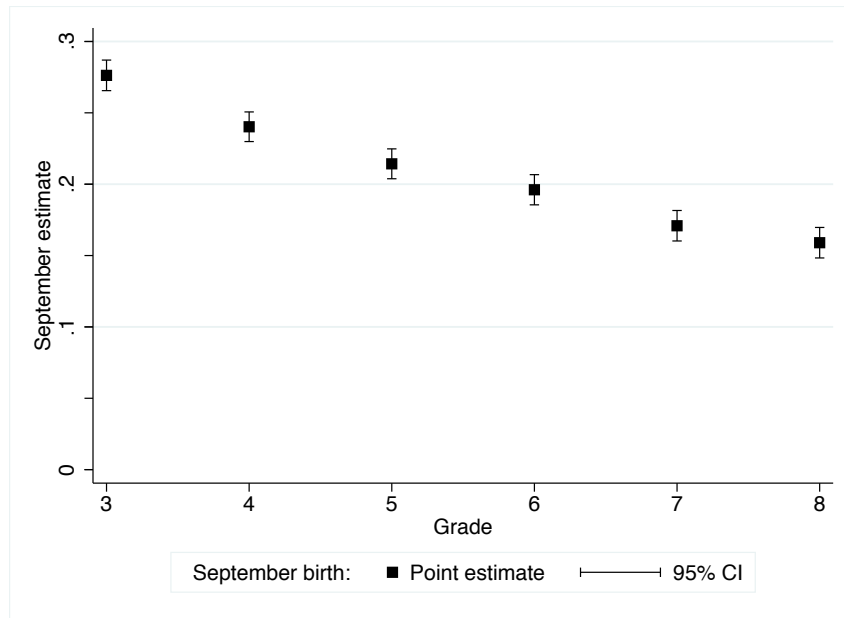
**Sprietsma**, “Effect of relative age in the first grade of primary school on long term scholastic results: International comparative evidence using PISA 2003,” *Education Economics*, 2010, *18* (1), 1–32.

**Tan, Poh Lin**, “The impact of school entry laws on female education and teenage fertility,” *Journal of Population Economics*, 2017, *30* (2), 503–536.

**Weil, Elizabeth**, “When should a kid start kindergarten,” New York Times June 3 2007.

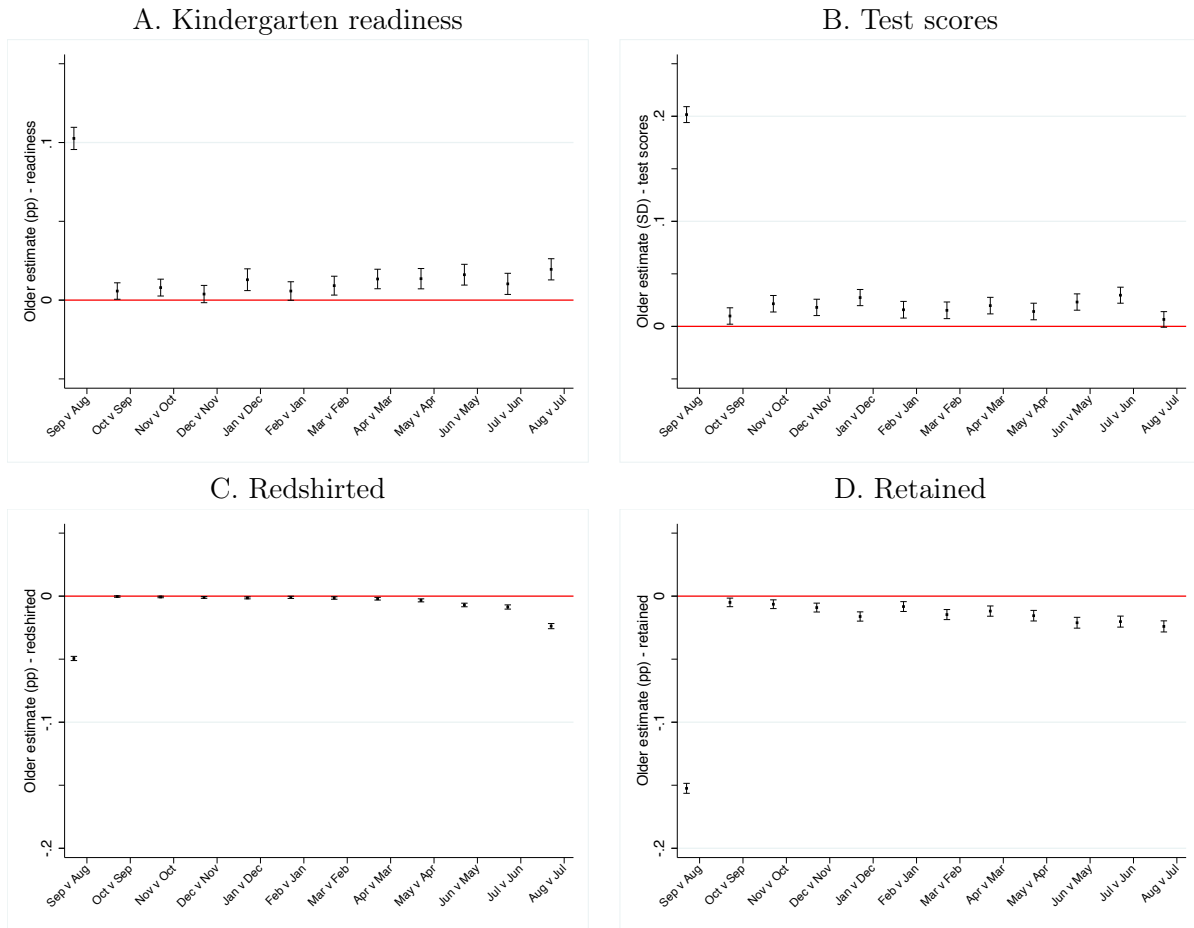
## Figures and Tables

Figure 1: Estimates of school starting age by grade



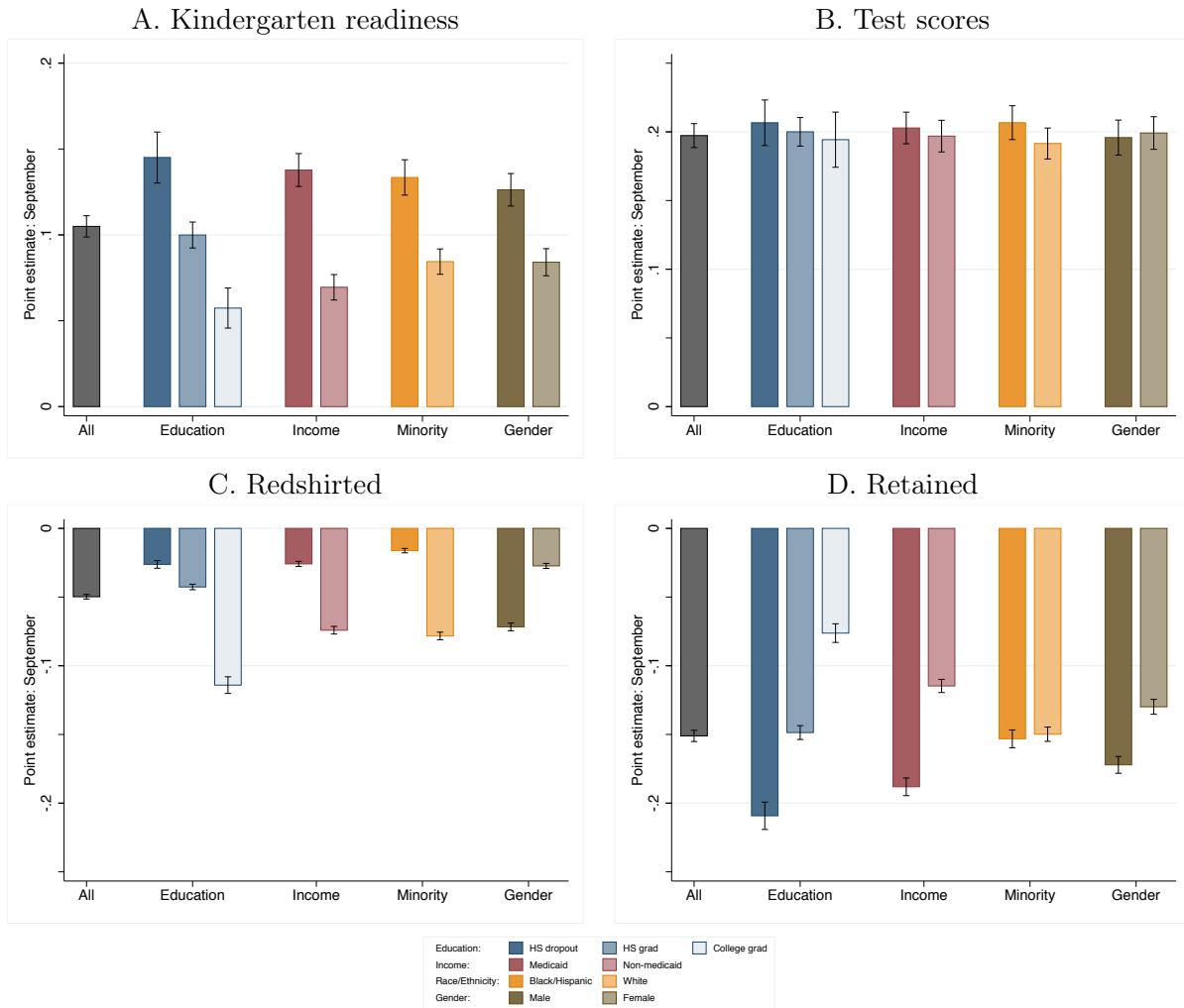
Note: Sample is based on all singleton births between 1994 and 2000. Each point estimate is based on regression of test scores in given grade (3 to 8) on an indicator for September birth, and a set of controls. Control variables include marital status at birth, maternal education indicators, indicator for medicaid paid birth, race and ethnicity indicators, indicator for gender, cohort dummies, log birth weight, gestational age, indicator for start of prenatal care in first trimester as well as indicators for congenital anomalies, abnormal conditions at birth and maternal health at birth. Heteroskedasticity robust standard errors and 95 percent confidence intervals.

Figure 2: Estimates of school starting age (month-by-month)



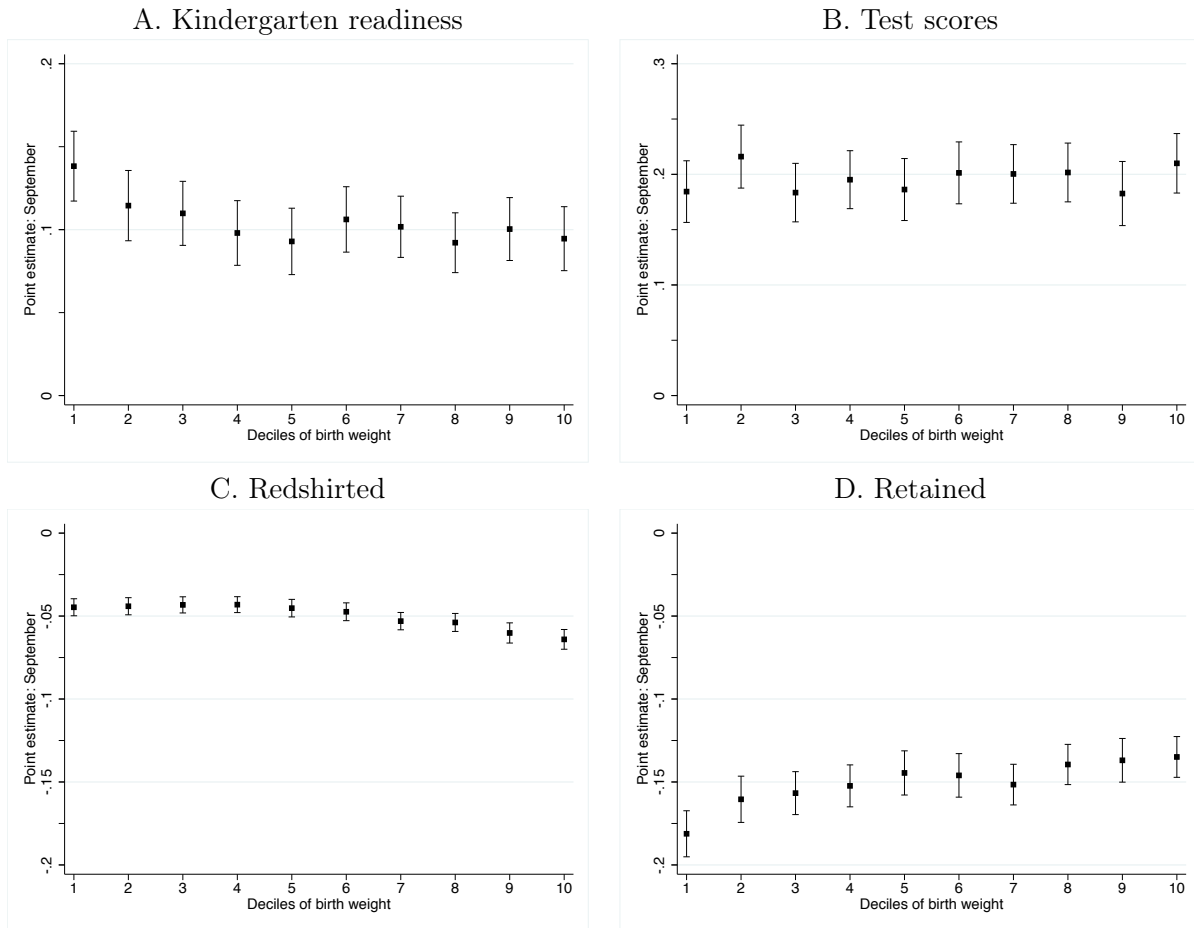
Note: Sample is based on all singleton births between 1994 and 2000. Each point estimate presents month-to-month comparison with 95 percent confidence intervals. Panel A presents results for kindergarten readiness, Panel B for pooled math and reading test scores in grades 3 to 8, Panel C for probability of being redshirted and Panel D for school retention. Kindergarten readiness excludes cohorts 1997 to 1999 due to missing data. Redshirting is defined as indicator variable that equals to one if a child has a higher than expected, based on date of birth, age at the time of first observation in school records in either kindergarten or grade one. School retention prior to grade three is defined as an indicator variable that equals to one if child is observed twice in the same grade. Control variables include marital status at birth, maternal education indicators, indicator for medicaid paid birth, race and ethnicity indicators, indicator for gender, cohort dummies, log birth weight, gestational age, indicator for starting prenatal care in first trimester as well as indicators for congenital anomalies, abnormal conditions at birth and maternal health problems. Heteroskedasticity robust standard errors in panels A, C and D and clustered at individual level in Panel B.

Figure 3: Heterogeneity by socioeconomic status and by gender (August vs. September)



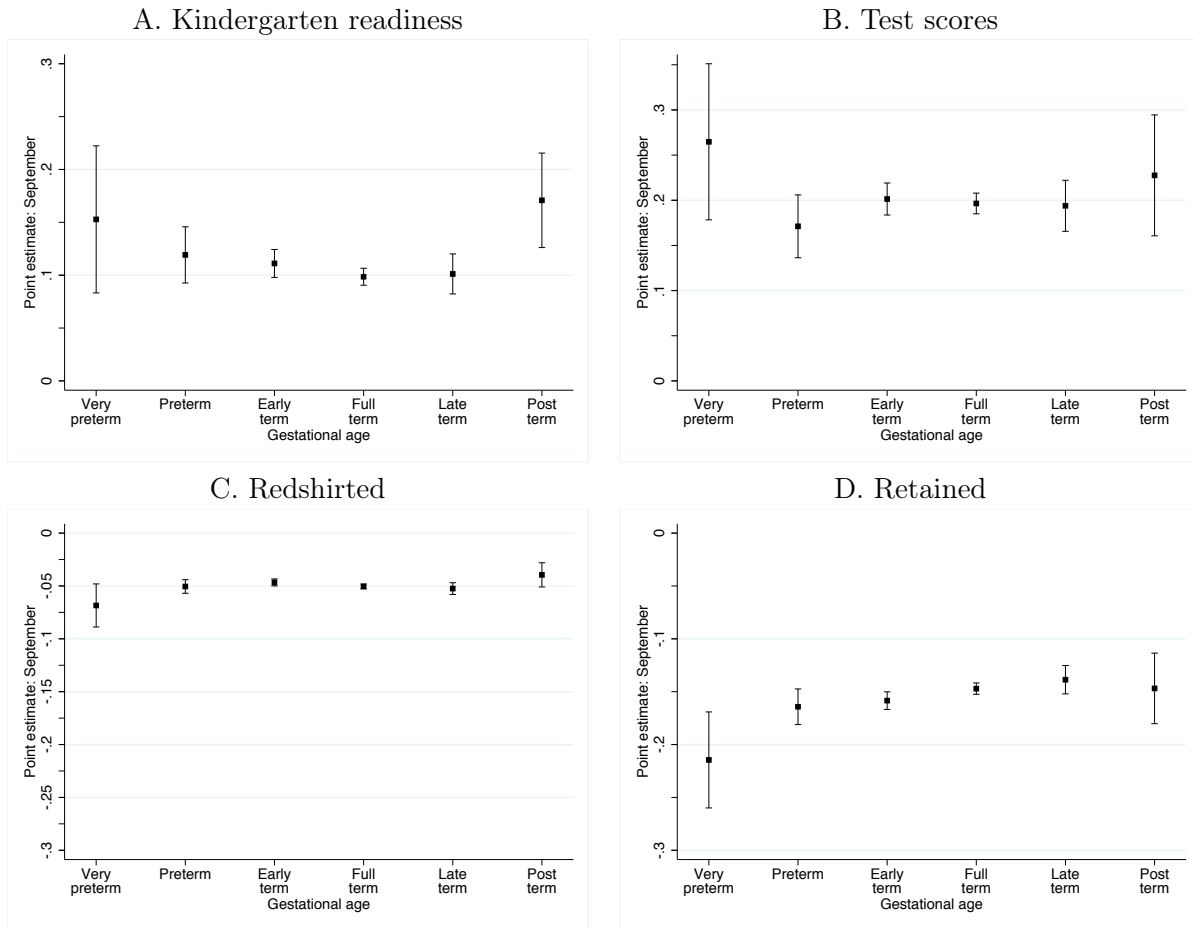
Note: Sample is based on all singleton births between 1994 and 2000. Each point estimate reflects August vs. September comparison with 95 percent confidence interval. Outcomes are: kindergarten readiness (Panel A), pooled math and reading test scores in grades 3 to 8 (Panel B), probability of being redshirted (Panel C) and school retention (Panel D). Black bars present average estimates akin to those in Figure 2; blue bars present heterogeneity by maternal education, maroon bars present heterogeneity by medicaid status which is proxy for income; orange bars present heterogeneity by race and ethnicity where minority is defined as either African-American or Hispanic; and olive bars present heterogeneity by gender. For definitions see Figure 2. No control variables are included. Heteroskedasticity robust standard errors for kindergarten readiness, being redshirted and retained while clustered at individual level for test scores.

Figure 4: Heterogeneity by birth weight (August vs. September)



Note: Sample is based on all singleton births between 1994 and 2000. Each point estimate reflects August vs. September comparison for each decile of birth weight with 95 percent confidence interval. Outcomes are: kindergarten readiness (Panel A), pooled math and reading test scores in grades 3 to 8 (Panel B), probability of being redshirted (Panel C) and school retention (Panel D). For definitions see Figure 2. No control variables are included. Heteroskedasticity robust standard errors for kindergarten readiness, being redshirted and retained while clustered at individual level for test scores.

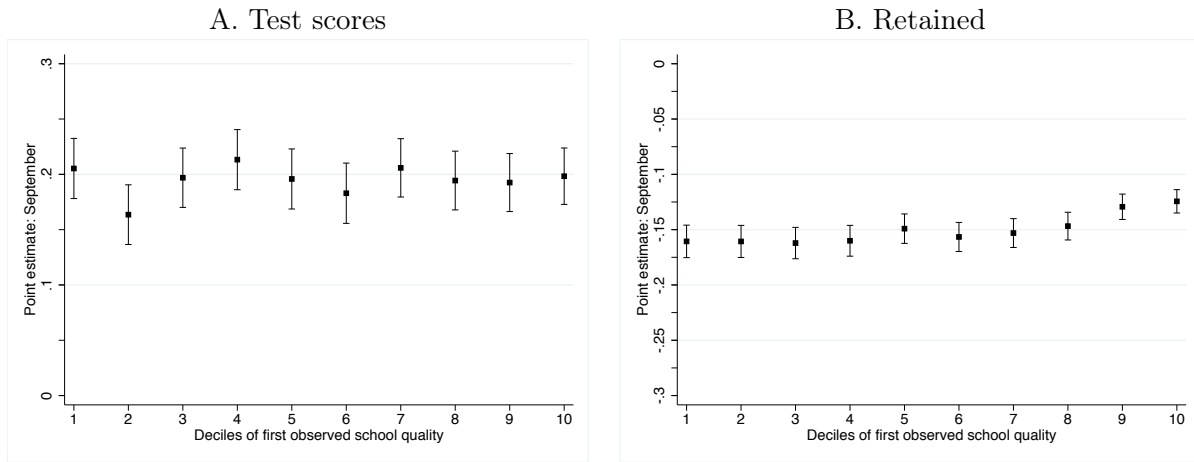
Figure 5: Heterogeneity by gestational age (August vs. September)



Note: Sample is based on all singleton births between 1994 and 2000. Each point estimate reflects August vs. September comparison for each gestational age group with 95 percent confidence interval. Gestational age groups are defined as follows: very preterm - below 32 weeks, preterm - 32 to 36 weeks, early term - 37 to 38 weeks, full term - 39 to 40 weeks, late term - 41 weeks, and post term - above 41 weeks. Outcomes are: kindergarten readiness (Panel A), pooled math and reading test scores in grades 3 to 8 (Panel B), probability of being redshirted (Panel C) and school retention (Panel D). For definitions see Figure 2. No control variables are included. Heteroskedasticity robust standard errors for kindergarten readiness, being redshirted and retained while clustered at individual level for test scores.

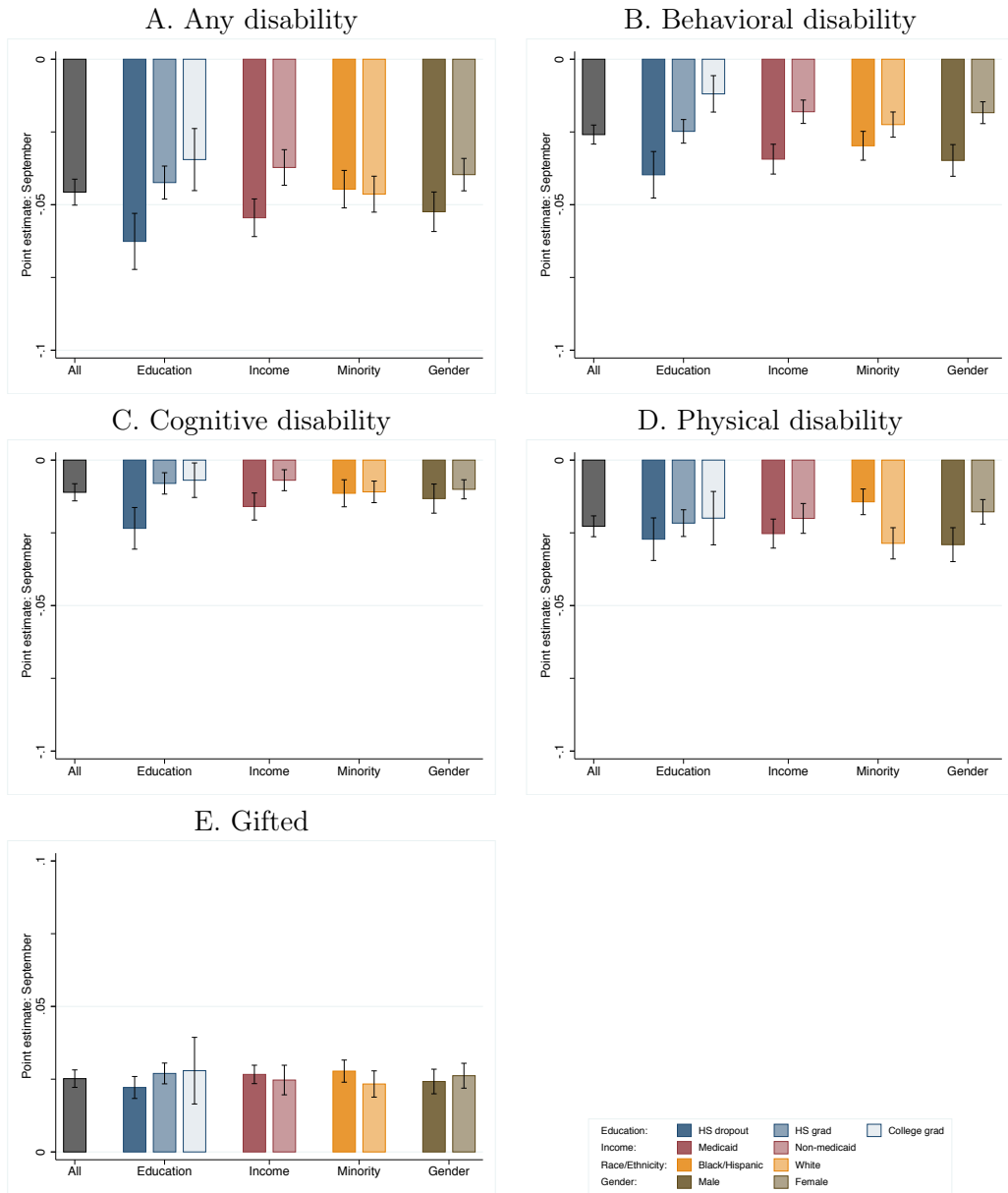


Figure 6: Heterogeneity by school quality (August vs. September)



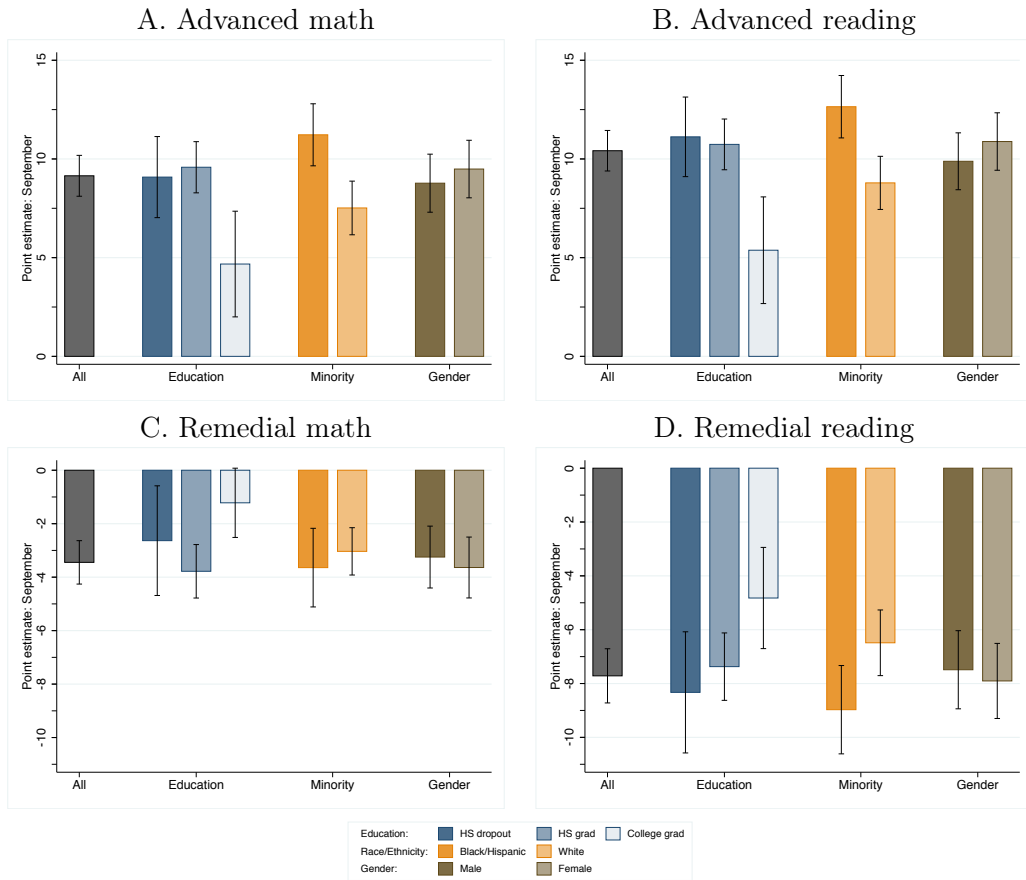
Note: Sample is based on all singleton births between 1994 and 2000. Each point estimate reflects August vs. September comparison for each decile of contemporaneous school quality with 95 percent confidence interval. Outcomes are: pooled math and reading test scores in grades 3 to 8 (Panel A) and school retention (Panel B). No control variables are included. Heteroskedasticity robust standard errors for being retained while clustered at individual level for test scores.

Figure 7: Effects of school starting age on disability - heterogeneity



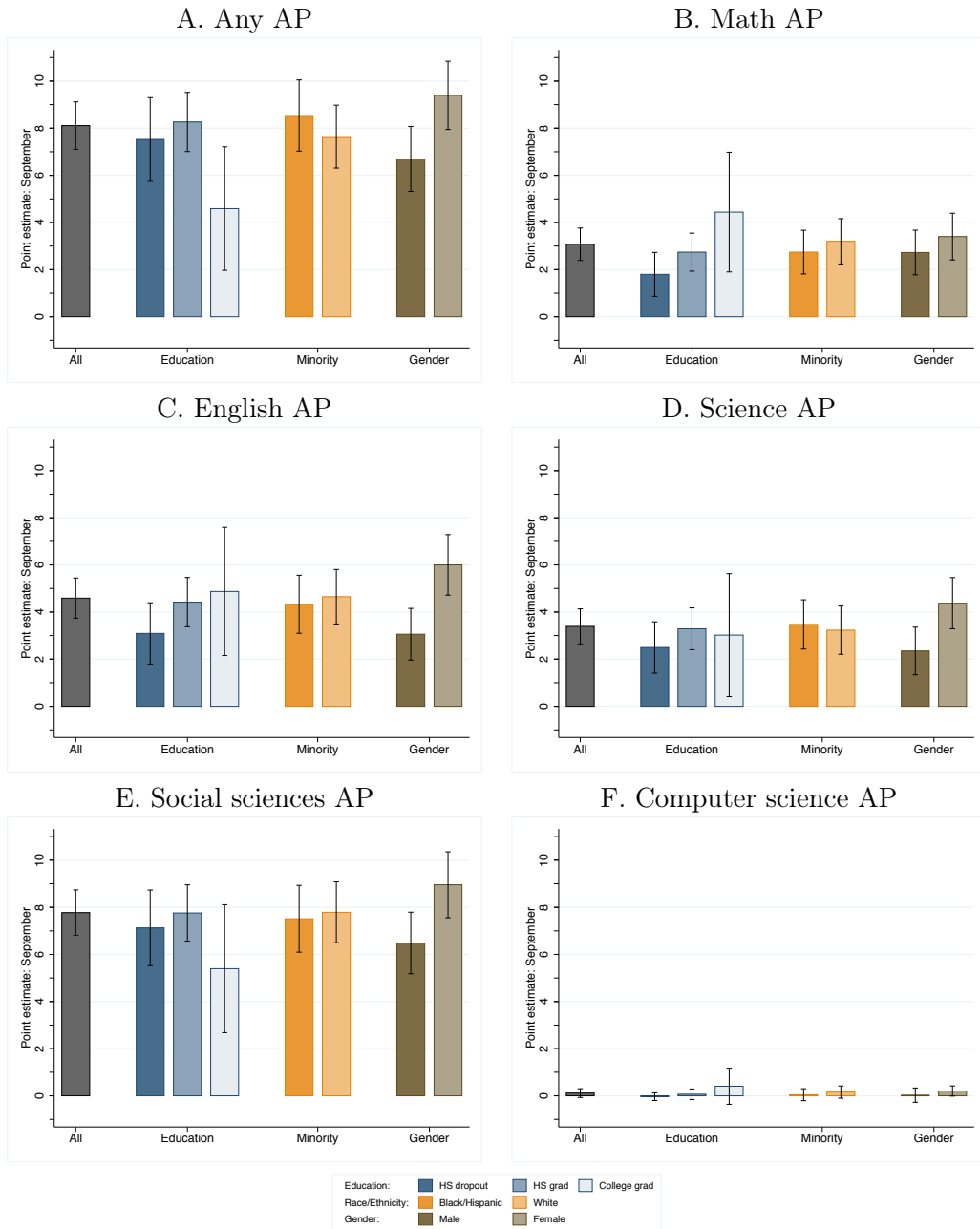
Note: Sample is based on all singleton births between 1994 and 2000. Each point estimate reflects August vs. September comparison with 95 percent confidence interval. Outcomes are diagnoses with: any disability (Panel A), behavioral disability (Panel B), cognitive disability (Panel C), physical disability (Panel D), and gifted status (Panel E). Black bars present average estimates; blue bars present heterogeneity by maternal education, maroon bars present heterogeneity by medicaid status which is proxy for income; orange bars present heterogeneity by race and ethnicity where minority is defined as either African-American or Hispanic; and olive bars present heterogeneity by gender. No control variables are included. Heteroskedasticity robust standard errors.

Figure 8: Effects of school starting age on middle school course enrollment - heterogeneity



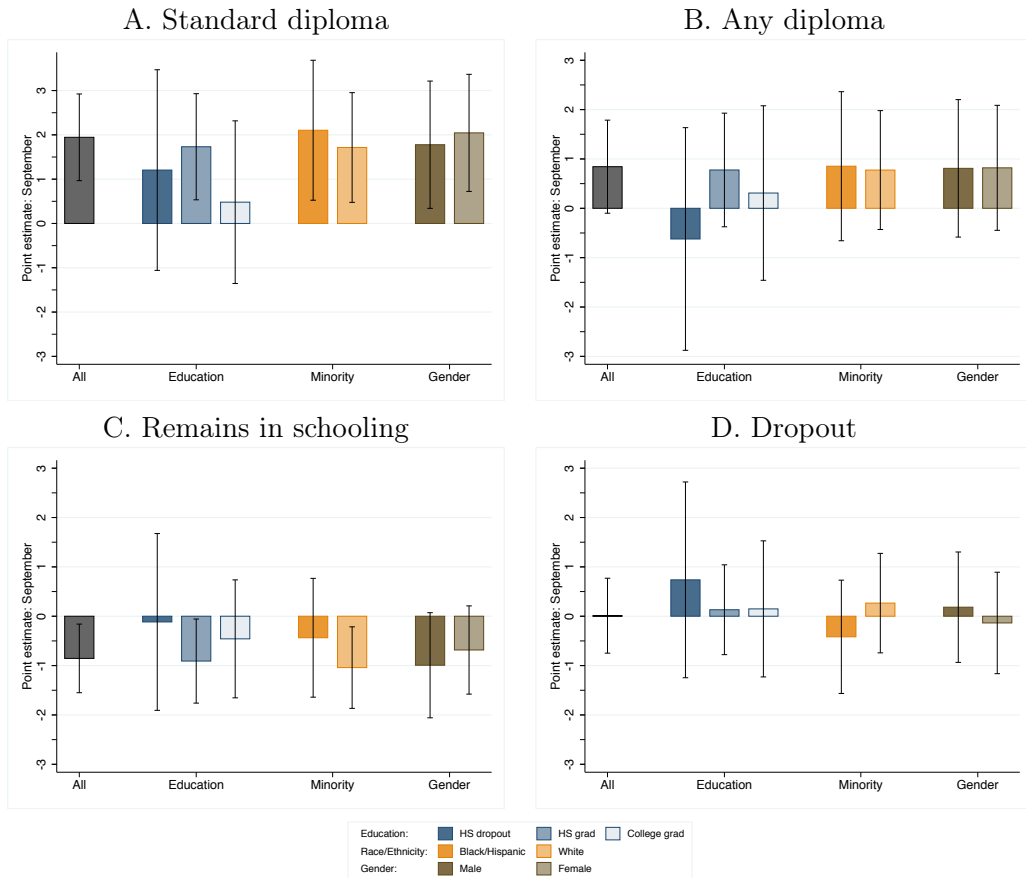
Note: Sample is based on all singleton births in 1992 and 1993. Each point estimate reflects August vs. September comparison with 95 percent confidence interval. Outcomes are enrollment in middle school in: advanced mathematics courses (Panel A), advanced reading courses (Panel B), remedial mathematics courses (Panel C) and remedial reading courses (Panel D). Black bars present average estimates; blue bars present heterogeneity by maternal education; orange bars present heterogeneity by race and ethnicity where minority is defined as either African-American or Hispanic; and olive bars present heterogeneity by gender. No control variables are included. Heteroskedasticity robust standard errors.

Figure 9: Effects of school starting age on high school course enrollment - heterogeneity



Note: Sample is based on all singleton births in 1992 and 1993. Each point estimate reflects August vs. September comparison with 95 percent confidence interval. Outcomes are enrollment in high school AP courses in: any AP course (Panel A), mathematics (Panel B), English (Panel C), science (Panel D), social sciences (Panel E) and computer science (Panel E). Black bars present average estimates; blue bars present heterogeneity by maternal education; orange bars present heterogeneity by race and ethnicity where minority is defined as either African-American or Hispanic; and olive bars present heterogeneity by gender. No control variables are included. Heteroskedasticity robust standard errors.

Figure 10: Effects of school starting age on graduation outcomes - heterogeneity



Note: Sample is based on all singleton births in 1992 and 1993. Each point estimate reflects August vs. September comparison with 95 percent confidence interval. Outcomes are: graduating high school with standard diploma (Panel A), graduating high school with any diploma (Panel B), remaining in schooling even though they should have graduated already (Panel C), and dropping out of high school (Panel D). Black bars present average estimates; blue bars present heterogeneity by maternal education; orange bars present heterogeneity by race and ethnicity where minority is defined as either African-American or Hispanic; and olive bars present heterogeneity by gender. No control variables are included. Heteroskedasticity robust standard errors.

Table 1: Effects of school starting age (August vs. September) - comparison of different econometric models

	(1)	(2)	(3)	(4)	(5)	(6)
	Grade 3 to 8 pooled test scores		Redshirted		Retained before third grade	
Panel A: Singletons (OLS)						
September birth	0.197*** (0.004)	0.202*** (0.004)	-0.050*** (0.001)	-0.049*** (0.001)	-0.151*** (0.002)	-0.152*** (0.002)
	[0.180 to 0.234][0.180 to 0.238]					
Mean of Y	0.063		0.028		0.202	
Observations	730,675				139,211	
N (children)			139,211			
Panel B: Siblings (OLS)						
September birth	0.216*** (0.025)	0.216*** (0.025)	-0.069*** (0.008)	-0.069*** (0.008)	-0.129*** (0.014)	-0.130*** (0.014)
	[0.212 to 0.223][0.216 to 0.224]					
Panel C: Siblings (FE)						
September birth	0.216*** (0.025)	0.218*** (0.025)	-0.069*** (0.008)	-0.069*** (0.008)	-0.129*** (0.014)	-0.131*** (0.014)
	[0.212 to 0.223][0.217 to 0.218]					
Mean of Y	0.146		0.037		0.177	
Observations	10,910				2,184	
N (sibling pairs)			1,092			
Panel D: Siblings with same parents (FE)						
September birth	0.223*** (0.029)	0.222*** (0.029)	-0.097*** (0.011)	-0.099*** (0.011)	-0.101*** (0.015)	-0.103*** (0.016)
	[0.216 to 0.234][0.220 to 0.227]					
Mean of Y	0.345		0.048		0.133	
Observations	7,476		1,470		1,470	
N (sibling pairs)	735		735		735	
Controls	X		X		X	

Note: Full sample is based on all singleton births between 1994 and 2000. All estimates come from August vs. September comparison. Samples are: universe of singletons (Panel A); siblings born one in each month (Panels B and C) and siblings born one in each month where the father is known and the same across the two births (Panel D). OLS regressions in Panels A and B while sibling fixed effects regressions in Panels C and D. Odd numbered columns do not include any controls while even numbered columns control for marital status at birth, maternal education indicators, indicator for medicaid paid birth, race and ethnicity indicators, indicator for gender, cohort dummies, log birth weight, gestational age, indicator for starting prenatal care in first trimester as well as indicators for congenital anomalies, abnormal conditions at birth and maternal health problems. In siblings models additional control is an indicator for second born. Standard errors clustered at individual level in columns (1) and (2) while heteroskedasticity robust standard errors in columns (3) to (6) in Panel A. Standard errors clustered at mother level in remaining panels (B to D). Square brackets in this Table present estimates from a bounding exercise that we perform to address selection into the estimation sample discussed in Section 2.1. In each case, we impute either the 5th or 95th percentile of test scores for children whom we observe without test scores. The imputed percentiles are computed separately for each year of birth, month of birth and grade in school so that we can account for the fact that later born children do not reach middle school grades by the end of our test scores data span. In particular, we do not impute test scores in grade 8 for children born in 2000 and in September of 1999; and we do not impute grade 7 for children born in September 2000. In panel A we impute scores for all children born in Florida who do not make it to our empirical sample while in panels B to D we do it conditionally on being observed in public school because only for this subsample we can identify siblings. The sample sizes for these bounding exercises are 1,231,791 in panel A; 16,350 in panels B and C; and 11,362 in panel D.

Table 2: Effects of school starting age (August vs. September) - disability and gifted status

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Any disability		Cognitive disability		Behavioral disability		Physical disability			Gifted status
September birth	-0.046*** (0.002)	-0.046*** (0.002)	-0.011*** (0.002)	-0.012*** (0.001)	-0.026*** (0.002)	-0.027*** (0.002)	-0.023*** (0.002)	-0.023*** (0.002)	0.025*** (0.002)	0.026*** (0.001)
Mean of Y	0.238		0.069		0.090		0.113		0.089	
Observations	139,211		113,991		116,521		119,549		139,211	
					Panel A: Singletons (OLS)					
September birth	-0.029* (0.017)	-0.030* (0.016)	-0.005 (0.011)	-0.006 (0.011)	-0.025** (0.012)	-0.024** (0.012)	-0.005 (0.014)	-0.004 (0.014)	0.023** (0.011)	0.024** (0.011)
Mean of Y	0.226		0.064		0.078		0.108		0.113	
Observations	2,184		1,532		1,584		1,692		2,184	
					Panel B: Siblings (OLS)					
September birth	-0.029* (0.017)	-0.032** (0.016)	-0.005 (0.011)	-0.005 (0.011)	-0.025** (0.012)	-0.026** (0.012)	-0.005 (0.014)	-0.008 (0.014)	0.023** (0.011)	0.021* (0.011)
Mean of Y	0.226		0.064		0.078		0.108		0.113	
Observations	2,184		1,532		1,584		1,692		2,184	
					Panel C: Siblings (FE)					
September birth	-0.046** (0.020)	-0.049** (0.020)	-0.014 (0.013)	-0.015 (0.013)	-0.034** (0.014)	-0.031** (0.014)	-0.010 (0.018)	-0.018 (0.018)	0.029* (0.015)	0.026* (0.015)
Mean of Y	0.219		0.067		0.067		0.122		0.152	
Observations	1,470		1,036		1,048		1,182		1,470	
Controls		X		X		X		X		X

Note: Sample is based on all singleton births between 1994 and 2000. All estimates come from August vs. September comparison based on specifications from columns (1) and (2) in Table 1. Samples in Panels A to D are equivalent to those used in Panels A to D in Table 1. Outcomes are: indicator for any disability (columns 1 and 2); indicator for cognitive disability (columns 3 and 4); indicator for behavioral disability (columns 5 and 6); indicator for physical disability (columns 7 and 8); and indicator for enrollment in gifted program (columns 9 and 10). Analyses in columns (3) to (8) compare each type of disability against population without any disabilities, and hence the sample size differs depending on disability considered. Heteroskedasticity robust standard errors in Panel A and standard errors clustered at mother level in Panels B to D.

Table 3: Effects of school starting age (August vs. September) - Middle and high school course selection

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Middle school					High school			
		Remedial courses					AP courses			
Advanced courses		Reading	Math	Reading	Any	Math	English	Science	Social Sci.	Comp. Sci.
Math										
September birth	9.147*** (0.528)	10.417*** (0.524)	-3.446*** (0.414)	-7.714*** (0.514)	8.109*** (0.514)	3.077*** (0.351)	4.591*** (0.435)	3.389*** (0.380)	7.775*** (0.491)	0.116 (0.094)
September birth	8.603*** (0.515)	9.988*** (0.511)	-2.934*** (0.399)	-6.911*** (0.484)	7.465*** (0.484)	2.713*** (0.337)	4.094*** (0.415)	3.022*** (0.365)	7.172*** (0.465)	0.097 (0.094)
Mean of Y	42.4	40.3	18.4	36.3	36.1	12.1	20.8	14.6	30.1	0.8
Observations	34,785	34,785	34,785	34,785	34,785	34,785	34,785	34,785	34,785	34,785

Panel A: no controls

Panel B: demographic and health controls

Note: Sample is based on all singleton births in 1992 and 1993. All estimates come from August vs. September comparisons. Panel A does not include any control variables while Panel B controls for maternal education dummies, marital status at the time of birth, race, ethnicity, nativity, gender, maternal age at the time of birth, cohort dummies, log birth weight, gestational age, indicator for start of prenatal care in first trimester as well as indicators for congenital anomalies, abnormal conditions at birth and maternal health at birth. Middle school course enrollment in columns (1) to (4) are: advanced mathematics, advanced reading, remedial mathematics, and remedial reading. High school AP course enrollment in columns (5) to (10) are: any course; mathematics, English, science, social sciences, and computer science. Heteroskedasticity robust standard errors.



Table 4: Effects of school starting age (August vs. September) - High school graduation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Graduated				Not-graduated			
	Standard diploma		Any diploma		Remains in schooling		Dropout	
September birth	1.945***	1.285***	0.843*	0.324	-0.854**	-0.519	0.011	0.195
	(0.499)	(0.474)	(0.481)	(0.462)	(0.355)	(0.349)	(0.387)	(0.378)
Mean of Y	68.2		72.1		12.5		15.4	
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	34,785							

Note: Sample is based on all singleton births in 1992 and 1993. All estimates come from August vs. September comparisons. Outcomes are: graduating high school with a standard diploma (columns 1 and 2); graduating high school with any diploma (columns 3 and 4); remaining in schooling even though they should have graduated already (columns 5 and 6), and dropping out of high school (columns 7 and 8). Odd numbered columns do not include any controls while even numbered columns control for maternal education dummies, marital status at the time of birth, race, ethnicity, nativity, gender, maternal age at the time of birth, cohort dummies, log birth weight, gestational age, indicator for start of prenatal care in first trimester as well as indicators for congenital anomalies, abnormal conditions at birth and maternal health at birth. Heteroskedasticity robust standard errors.

Table 5: Interaction between school policies and effects of school starting age

	(1)	(2)	(3)		(4)	(5)	(6)
	Univariate	Multivariable	% Yes		Univariate	Multivariable	% Yes
A. Does this school sponsor any of the following before-school or after-school programs?				C. Does this school sponsor?			
Child care programs	0.002 (0.015)	-0.006 (0.015)	83.4	Summer school	-0.003 (0.012)	-0.018 (0.013)	67.9
Recreational programs	0.014 (0.011)	0.003 (0.012)	42.5	Year round classes	-0.031 (0.039)	-0.033 (0.039)	2.1
Academic enrichment	0.026** (0.011)	0.016 (0.012)	50.0	Extended school year	-0.014 (0.015)	-0.008 (0.015)	17.0
Remedial/tutoring program	0.019 (0.016)	0.004 (0.018)	85.3	Saturday school	0.013 (0.019)	-0.010 (0.025)	9.6
B. Does this school structure schedules and staff in any of the following ways?				D. What is the average number of students for a regular class?			
Block scheduling	0.029** (0.011)	0.023* (0.012)	36.5	Class size <sup>^</sup>	0.002** (0.001)	0.001* (0.001)	23.6
Common preparation periods	-0.001 (0.017)	-0.007 (0.018)	89.0	E. What special measures, if any, does this school take to try to improve the performance of low performing students?			
Subject specialist teacher	0.012 (0.011)	0.005 (0.012)	54.8	Require grade retention	0.004 (0.013)	-0.009 (0.014)	75.6
Organize teachers into teams	-0.014 (0.029)	-0.005 (0.029)	96.4	Require summer school for grade advancement	0.025** (0.011)	0.023* (0.012)	58.9
Looping	-0.007 (0.011)	-0.005 (0.012)	44.7	Require school supplemental instruction	-0.014 (0.014)	-0.022 (0.014)	79.5
Multi age classrooms	0.006 (0.012)	0.008 (0.012)	32.3	Require Saturday classes	0.043 (0.028)	0.039 (0.037)	4.4
				Require before/after school tutoring	0.017 (0.011)	0.009 (0.013)	48.3
Mean of Y				0.121			
# children				83,510			

Note: Sample is based on all August and September singleton births between 1994 and 2000. It is further restricted to individuals attending grade 1 in schools for which we observe complete information on all policies in question and observed with test scores in grade 3. Outcome variable is test scores in grade 3. We display coefficient on the interaction between indicator for September birth and indicator for school using a given policy, and regressions also control for both of those indicators. All regressions further control for log birth weight, gestational age, indicators for prenatal case started in first trimester, congenital anomalies, abnormal conditions at birth and maternal health problems as well as indicators for birth cohort, maternal education, medicaid birth, race and ethnicity and child's gender. Columns (1) and (4) only include a single interaction at a time while columns (2) and (5) include all interactions together in one regression. Columns (3) and (6) present means for policy use (<sup>^</sup> marks average class size in column 6). Heteroskedasticity robust standard errors.

Table 6: Effects of school starting age (August vs. September) - utilizing regional variation in scores, redshirting and retentions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Full sample		Maternal education		Heterogeneity		Minority	
	HS dropout	HS grad	College grad	Yes	No	Yes	No	
September birth*% redshirted	-0.225* (0.114)	-0.405* (0.238)	-0.265 (0.262)	-0.159 (0.324)	-0.269* (0.160)	-0.139 (0.208)	-0.035 (0.144)	-0.073 (0.114)
% redshirted	0.078*** (0.027)	0.084 (0.062)	0.061 (0.042)	0.071*** (0.024)	0.028 (0.055)	0.054* (0.028)	-0.076 (0.052)	0.061* (0.033)
September birth	0.699*** (0.069)	0.275*** (0.082)	0.315*** (0.097)	0.384*** (0.128)	0.414*** (0.051)	0.489*** (0.091)	0.483*** (0.046)	0.502*** (0.050)
September birth*% retained	-0.188*** (0.070)	-0.100 (0.068)	-0.174** (0.080)	-0.234 (0.149)	-0.135* (0.070)	-0.299*** (0.085)	-0.121*** (0.035)	-0.243*** (0.080)
% retained	-0.183*** (0.058)	-0.104* (0.060)	-0.049 (0.043)	-0.294*** (0.092)	-0.111** (0.053)	-0.145** (0.058)	-0.141*** (0.043)	-0.180*** (0.057)
September birth	0.433*** (0.068)	0.303*** (0.075)	0.258*** (0.045)	0.069 (0.117)	0.373*** (0.064)	0.164*** (0.056)	0.349*** (0.044)	0.142*** (0.050)
N	910		449			712		624
N (districts)	65		41			62		54

Note: Sample is based on all singleton births between 1994 and 2000. All regressions are run on aggregated means data where the variables are collapsed at year of birth-school-district-September birth level. Analysis is based on 65 school districts and we have to exclude two small school districts because they do not record place of residence at birth for years 1994 and 1995. These two districts constitute 1.5 percent of the full population of births in years 1996 to 2000. Each regression controls for school district and cohort fixed effects and weights the estimates by number of children in cohort-district-September birth cells. Panel A presents results for the relationship between test scores and redshirting while Panel B between test scores and early retention. Each panel presents estimates on indicator for September births, fraction of children experiencing given explanatory variable (redshirting or retention) and the interaction between these two variables. Column (1) presents results for full sample: 65 districts, 7 years and 2 birth months. Columns (2) to (4) present the analysis split additionally by maternal education (cell is district, year of birth, September birth and three maternal education groups). Columns (5) and (6) present the analysis split by maternal Medicaid status (cell is district, year of birth, September birth and two Medicaid groups). Columns (7) and (8) present the analysis split by racial/ethnic minority status (cell is district, year of birth, September birth and two maternal race/ethnicity groups). In the heterogeneity analysis we require at least five observations in each cell and all three/two heterogeneity dimensions in each cell. This yields unbalanced repeated cross-section in the heterogeneity analyses. Standard errors are clustered at school district level.

## Appendix

### A1. Florida school survey

We utilize following questions in our analysis in Table 5:

1. Does this school sponsor any of the following before-school or after-school programs? (yes/no)
  - (a) child care programs
  - (b) recreational programs
  - (c) academic enrichment programs
  - (d) remedial/tutoring programs
2. Does this school structure schedules and staff in any of the following ways? (yes/no)
  - (a) block scheduling
  - (b) common preparation periods
  - (c) subject specialist teacher
  - (d) organize teachers into teams
  - (e) looping
  - (f) multi age classrooms
3. Does this school sponsor? (yes/no)
  - (a) summer school
  - (b) year round classes
  - (c) extended school year
  - (d) Saturday school
4. What is the average number of students for a regular class? (number; grade specific)
5. What special measures, if any, does this school take to try to improve the performance of low performing students?
  - (a) require grade retention
  - (b) require summer school for grade advancement
  - (c) require school supplemental instruction
  - (d) require Saturday classes
  - (e) require before/after school tutoring

For questions 1, 2, 3 and 5 we code indicator equal to one if principal responded affirmatively in the first survey year. In question 4 we chose the number of students reported in grade one. We discard all schools with missing observations in any of the questions.

## A2. Tables

Table A1: Descriptive statistics: demographic characteristics of mothers and children

	(1)	(2)	(3)	(4) (5)		(6)	(7)	(8)
	All births	All	Singletons	sample used in analysis		Sibling sample used in analysis		
			All	August	September	All	August	September
% African-American	21.9	22.4	25.8	25.7	25.9	24.2	24.2	24.2
% Hispanic	22.7	23.0	24.1	23.9	24.3	23.6	23.6	23.6
% immigrant	23.0	23.3	23.1	22.8	23.4	20.0	20.0	20.0
% HS dropout	20.6	20.6	23.8	23.6	24.0	25.0	25.0	25.0
% HS grad	59.0	59.3	61.0	61.0	61.0	55.1	55.1	55.1
% college grad	20.5	20.1	15.2	15.4	15.0	19.9	19.9	19.9
% married	65.6	65.2	60.7	61.0	60.3	63.4	63.1	63.7
% Medicaid birth	44.4	45.1	50.8	50.5	51.1	50.4	50.4	50.4
% male	51.2	51.1	50.6	50.5	50.8	51.6	51.5	51.6
% mom health problems	23.7	23.7	24.3	24.4	24.3	23.2	23.0	23.4
Maternal age	27.1	27.1	26.6	26.6	26.6	24.8	24.8	24.8
Birth weight	3343	3341	3328	3325	3330	3318	3318	3319
% September	8.8	50.0	48.8	0.0	100.0	50.0	0.0	100.0
N	1,220,803	215,971	139,211	71,214	67,997	2,184	1,092	1,092

Note: Sample is based on all singleton births between 1994 and 2000. Table A1 present means and sample sizes for eight different samples. Column (1) includes all births between 1994 and 2000 with complete demographic information; column (2) presents a subset of these births from August and September. Columns (3) to (5) present information for children used in the singletons empirical analysis while columns (6) to (8) are restricted to sample of siblings used in the sibling fixed effects empirical analysis. Columns (3) and (6) present descriptives for pooled August and September births while columns (4), (5), (7) and (8) present it for each month and sample separately.

Table A2: Effects of school starting age (August vs. September) - separate estimates for mathematics and reading

	(1)	(2)	(3)	(4)
	Grade 3 to 8 pooled test scores in math		Grade 3 to 8 pooled test scores in reading	
Panel A: Singletons (OLS)				
September birth	0.186*** (0.005)	0.190*** (0.004)	0.208*** (0.005)	0.213*** (0.004)
Mean of Y	0.062		0.065	
Observations	722,642		728,913	
Number of children	139,038		139,188	
Panel B: Siblings (OLS)				
September birth	0.195*** (0.026)	0.195*** (0.027)	0.239*** (0.027)	0.238*** (0.027)
Mean of Y	10,758		10,874	
Panel C: Siblings (FE)				
September birth	0.195*** (0.026)	0.199*** (0.026)	0.239*** (0.027)	0.239*** (0.026)
Mean of Y	0.163		0.133	
Observations	10,758		10,874	
Number of sibling pairs	1,092		1,092	
Panel D: Siblings with same parents (FE)				
September birth	0.209*** (0.031)	0.208*** (0.031)	0.237*** (0.031)	0.238*** (0.031)
Mean of Y	0.364		0.326	
Observations	7,392		7,456	
Number of sibling pairs	735		735	
Controls	X		X	

Note: This table replicates analysis from columns (1) and (2) of Table 1 separately for mathematics (columns 1 and 2) and reading (columns 3 and 4) test scores. Standard errors clustered at individual level in Panel A and at mother level in Panels B to D.

Table A3: Effects of school starting age (August vs. September) - comparison of different econometric models, continued

	(1)	(2)	(3)	(4)	(5)	(6)
	Grade 3 to 8 pooled test scores					
	OLS		Reduced form		Instrumental variables	
	(age at test)		(September birth)		(age at test)	
Point estimate	-0.040***	-0.030***	0.197***	0.202***	0.307***	0.323***
	(0.001)	(0.000)	(0.004)	(0.004)	(0.007)	(0.007)
First-stage	N/A		N/A		0.642***	0.624***
					(0.003)	(0.003)
Mean of Y			0.063			
Observations			730,675			
# children			139,211			
Controls	X		X		X	

Note: This table is based on sample and analysis from columns (1) and (2) in Panel A of Table 1. Panel A regresses test scores on age at the time of test. Panel B regresses test scores on indicator for September birth. Analyses in Panel B replicate results from Panel A of Table 1 for comparison. Panel C presents 2SLS estimates where in the first-stage we regress age at the time of test on September birth while in the second-stage we regress test scores on predicted age at the time of test. Age at the time of test is defined as age in months in March of a given school year. FCAT test is administered in late February to mid-March. Standard errors clustered at individual level.

Table A4: Effects of school starting age (August vs. September) - selection into public schools

	(1)	(2)	(3)	(4)	(5)	(6)
	Singletons			Sibling FE		
	P(matched to public schools)		P(observed with 3rd grade test scores   matched to public schools)			
September birth	-0.019***	-0.020***	0.006***	0.005***	0.005	0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.010)	(0.010)
Mean of Y	0.807		0.818		0.833	
Observations	215,971		174,439		2,952	
Controls	X		X		X	

Note: Sample is based on all singleton births between 1994 and 2000. All estimates come from August vs. September comparison. The dependent variable in columns 1 and 2 is probability of being matched between birth records and public school records. The dependent variable in columns 3 to 6 is probability of being observed with third grade test score conditional on being matched between birth and public school records. Samples are: universe of singleton births (columns 1 and 2); universe of singleton births matched to public school records (columns 3 and 4); and subsample of siblings born one in each month (columns 5 and 6). Cross-sectional regressions in columns 1 to 4 and sibling fixed effects regressions in columns 5 and 6. Columns 1, 3 and 5 do not include any controls; columns 2, 4 and 6 control for maternal education, marital status at birth, Medicaid birth, race, ethnicity, child's gender, cohort dummies, log birth weight, gestational age, indicator for start of prenatal care in first trimester as well as indicators for congenital anomalies, abnormal conditions at birth and maternal health at birth. Column 6 further includes indicator for second born. Robust standard errors in columns 1 to 4 and clustered at family level in columns 5 and 6.

Table A5: Effects of school starting age (August vs. September) - differential effects for boys by maternal socioeconomic characteristics

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	HS dropout	Maternal education HS grad	College grad	Income Medicaid	Non-medicaid	Race/Ethnicity Black/Hispanic	White
Panel A: Kindergarten readiness							
September effect for boys	0.149*** (0.011)	0.126*** (0.006)	0.077*** (0.010)	0.155*** (0.007)	0.093*** (0.006)	0.149*** (0.008)	0.111*** (0.006)
September effect for girls	0.139*** (0.010)	0.074*** (0.005)	0.041*** (0.007)	0.119*** (0.006)	0.047*** (0.005)	0.119*** (0.007)	0.058*** (0.005)
p-value difference	0.540	p<0.001	0.003	p<0.001	p<0.001	0.004	p<0.001
Observations	12,532	31,665	7,247	26,764	24,680	22,977	28,467
Panel B: Test scores							
September effect for boys	0.212*** (0.013)	0.198*** (0.008)	0.163*** (0.015)	0.209*** (0.009)	0.182*** (0.008)	0.207*** (0.009)	0.187*** (0.008)
September effect for girls	0.201*** (0.011)	0.203*** (0.007)	0.229*** (0.014)	0.197*** (0.008)	0.213*** (0.008)	0.207*** (0.008)	0.197*** (0.008)
p-value difference	0.515	0.635	0.001	0.299	0.007	0.975	0.385
Observations	172,587	447,011	111,077	368,859	361,816	338,780	391,895
Number of individuals	33,132	84,946	21,133	70,701	68,510	64,342	74,869
Panel C: Redshirted							
September effect for boys	-0.030*** (0.002)	-0.063*** (0.002)	-0.169*** (0.005)	-0.033*** (0.002)	-0.110*** (0.002)	-0.020*** (0.001)	-0.114*** (0.002)
September effect for girls	-0.022*** (0.002)	-0.022*** (0.001)	-0.057*** (0.003)	-0.018*** (0.001)	-0.037*** (0.001)	-0.012*** (0.001)	-0.041*** (0.002)
p-value difference	0.002	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001
Observations	33,132	84,946	21,133	70,701	68,510	64,342	74,869
Panel D: Retained							
September effect for boys	-0.214*** (0.007)	-0.175*** (0.004)	-0.097*** (0.005)	-0.206*** (0.005)	-0.139*** (0.004)	-0.166*** (0.005)	-0.178*** (0.004)
September effect for girls	-0.203*** (0.007)	-0.122*** (0.003)	-0.055*** (0.004)	-0.170*** (0.004)	-0.090*** (0.003)	-0.141*** (0.004)	-0.121*** (0.003)
p-value difference	0.256	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001
Observations	33,132	84,946	21,133	70,701	68,510	64,342	74,869

Note: Sample is based on all singleton births between 1994 and 2000. For each sample and outcome we present two estimates on being born in September separately for males and females. The p-values reported below each estimates pair test statistical equality of the two coefficients. Columns (1) to (3) present heterogeneity by maternal education, columns (4) and (5) present heterogeneity by medicaid status which is proxy for income, and columns (6) and (7) present heterogeneity by race and ethnicity. Outcomes are: kindergarten readiness (Panel A), pooled math and reading test scores in grades 3 to 8 (Panel B), probability of being redshirted (Panel C) and school retention (Panel D). Kindergarten readiness excludes cohorts 1997 to 1999 due to missing data. Redshirting is defined as indicator variable that equals to one if a child has a higher than expected, based on date of birth, age at the time of first observation in school records in either kindergarten or grade one. School retention prior to grade three is defined as an indicator variable that equals to one if child is observed twice in the same grade. No control variables are included. Heteroskedasticity robust standard errors for being redshirted and retained while clustered at individual level for test scores.