Forced to Redshirt: Quasi-Experimental Impacts of Delayed Kindergarten Entry

Jade M. Jenkins
University of California, Irvine

C. Kevin Fortner
Georgia State University

We provide causal estimates of the effects of delayed kindergarten entry on achievement outcomes by exploiting a policy change in the birthdate enrollment cutoff in North Carolina that forced children born in a six-week window to redshirt. Using multiple peer group comparisons, we identify impacts on achievement and gifted or disability identifications in third through fifth grades. Delayed entry provides small benefits to students' math and reading achievement, and reduced identification of a disability; these impacts operate through cohort position and age advantages, and not from hold-out year experiences. Redshirting differentially benefitted low-income students, but further disadvantaged non-white students.

Forced to Redshirt: Quasi-Experimental Impacts of Delayed Kindergarten Entry

Jade Marcus Jenkins
C. Kevin Fortner

Abstract
We provide causal estimates of the effects of delayed kindergarten entry on achievement
outcomes by exploiting a policy change in the birthdate enrollment cutoff in North Carolina that
forced children born in a six-week window to redshirt. Using multiple peer group comparisons,
we identify impacts on achievement and gifted or disability identifications in third through fifth
grades. Delayed entry provides small benefits to students' math and reading achievement, and
reduced identification of a disability; these impacts operate through cohort position and age
advantages, and not from hold-out year experiences. Redshirting differentially benefitted low-
income students, but further disadvantaged non-white students.

I.   Introduction

A new concern in parents' human capital investment decisions is whether to enroll their child in kindergarten on time, when they are first age-eligible, or to delay kindergarten entry (Paul, 2010). The latter behavior is otherwise known as kindergarten "redshirting", a term borrowed from collegiate sports, whereby children sit out their expected entry year to better mature and gain skills, and begin the following year as an old-for-grade student. Redshirting is now more than a trend, with national rates estimated between 4 to 14 percent (Deming & Dynarski, 2008). Influencing these decisions is some early evidence of the positive effects from entering school at an older age for some academic and socioemotional outcomes (Datar, 2006b; Stipek, 2002 provides a comprehensive review). In light of this evidence, states have also shifted their birthdate cutoffs for school entry to earlier in the school year (e.g., from December to September) with hopes that older cohorts will provide boosts to statewide test scores.

Yet results from the more rigorous literature that followed this early work on the effects of delaying school entry on children's outcomes are equivocal. Causal estimates of the effects of delayed kindergarten entry are difficult to obtain because redshirting is a clearly selected behavior. Parents redshirt their children in response to local social mores, a desire to improve their children's readiness for school, and out of concern for their children's development (e.g., perceived developmental delay), resulting in both positive and negative selection into redshirting (Fortner & Jenkins, 2017). Furthermore, redshirting is costly in terms of an additional year of child care or parents' foregone earnings from home care, and in terms of the instructional and administrative adjustments districts and teachers must make when policy changes the age composition of their classrooms. Thus, policymakers and parents have made these decisions largely in the dark as to whether delaying entrance into kindergarten will produce meaningful improvements in achievement that outweigh such costs.

Our paper addresses this gap by providing the first causal estimates of the effects of redshirting on achievement outcomes by exploiting an exogenous policy change in the birthdate enrollment cutoff for public schools in North Carolina (NC). NC moved the birthdate cutoff from Oct 15th to September 1st for children kindergarten in the 2009-2010 school year. This change required children born during these six weeks to delay kindergarten entry to the following year. We compare the outcomes of these "forced" redshirted children with the outcomes of their peers using census-level administrative data that include exact birthdates to identify the impact of delaying kindergarten entry on students' achievement and important curricular determinations of giftedness or having a disability in elementary school. We look at outcomes across 3rd through 5th grades to see if these impacts fade out in the intermediate term.

Exploiting both the policy change and the use of birthdates to assign students to cohorts allows us to estimate the causal impacts of redshirting *and* to assess the causal mechanisms behind these impacts. This is important because redshirting conflates three separate "treatments", each of which may impact achievement outcomes: (1) *Aging,* whereby a full year of maturation may increase children's attention spans and their tolerance for seated instruction, or improve their behavior (Meisels, 1999); (2) *Cohort position,* whereby shifting from the youngest to the oldest of their class peers give children relative age and size advantages over their peers (i.e., Gladwell, 2008); and (3) *Hold-out year experiences*, like high quality early childhood education, that provide children with skills that make them more successful in school (Elder & Lubotsky, 2009). Through a multiple comparison method approach, we are able to triangulate which of these three mechanisms is driving the estimated impacts of delayed entry. Thus, understanding *why* delayed entry may affect student outcomes is another important contribution of our study.

The estimates generated from our unique approach reveal an interesting story; overall, delayed kindergarten entry is correlated with small test score and academic designation

advantages. Forced redshirts get a math and reading test score boost between 0.04 and 0.08 standard deviations (SD) compared with the relatively younger students in their newly assigned cohort. They are also less likely to be identified as disabled. These effects persist between third and fifth grade. However, when we remove the effects of aging (mechanism 1, above) from the estimation, most of these advantages disappear, save for some small effects on reading achievement. When we go on to also remove the effect of cohort position from the estimate of redshirting (mechanism 2), there are no longer any statistically significant impacts on achievement, but some suggestive evidence of a small, one percentage point, decrease in the probability of being identified as disabled. This suggests that the overall impacts of redshirting— for those students impacted by the change in enrollment cutoff—are largely driven by aging and cohort position effects and not from hold-out year experiences.

Because we use population-level data from a large and diverse state, another key contribution of our study is our ability to test for potential heterogeneous impacts of delaying kindergarten entry across subgroups defined by sex, race/ethnicity, and income. This is important because it allows us to evaluate the impact of a birthdate cutoff policy change on students who are traditionally marginalized or disadvantaged in public education. Children from low-income families do worse on various measures of school readiness and academic ability at kindergarten entry compared with their higher income peers (Duncan & Magnuson, 2013; Fryer & Levitt, 2006). The cost of an additional year of child care for low-income families who are forced to redshirt may decrease educational opportunities for such children if they spend the year in low-quality care (Datar estimates that the additional childcare cost burden to parents from changes in entrance age policies is approximately $115 million (2006b)). [i] Furthermore, children who are white, male, and from higher income families are also more likely to redshirt; the combination of higher incidence rates and differential impacts from redshirting (in favor of advantaged groups)

could exacerbate preexisting achievement differences between racial and ethnic groups or socioeconomic groups, and between boys and girls.

We find that the policy-induced delayed entry impacts were stronger for economically disadvantaged children (measured as free and reduced-price lunch eligibility) across several outcomes. Low-income students benefitted nearly twice as much in terms of math and reading achievement increases and reductions in disability identifications relative to their higher income peers as a result of delayed entry. However, the birthdate policy shift *widened* gaps between white and Hispanic students in reading and math achievement and identification as gifted. Delaying entry provided black students with a gain in reading scores relative to white students, but not in math scores or academic designations. We did not find systematic differences in policy impacts by sex. In terms of policy levers, our findings suggest that changing birthdate entry cutoffs do not yield pareto improvements, and could indeed be harmful in the pursuit of equity in public education.

Our paper proceeds with a review of the literature on delayed entry, relative age and human capital outcomes, followed by a description of the data, the NC policy context, analyses, and results. Our final section concludes and discusses implications.

II.   Background

The idea that older children are more ready or able to benefit from schooling comes from developmental theory suggesting that biological maturation provides an added advantage for classroom-based learning (Meisels, 1999; Morrison, Alberts, & Griffith, 1997). With age, children may gain competencies that will help them succeed in school. Parents may fear that children who may be too immature for a classroom at kindergarten would be perpetually disadvantaged over the course of their schooling through a cascade of negative outcomes

(anxiety, underachievement, grade retention). This conventional wisdom, along with accountability pressure, led to a policymaking trend whereby states shifted school entrance age cutoffs for kindergarten to earlier in the year, requiring that children are at least five years old at the start of their kindergarten school year (e.g. moving the date from December to September). In 1975, six states set cutoffs at September 14 or earlier, compared with 33 states in 2014 (Education Commission of the States, 2014; Elder & Lubotsky, 2009). However, the early research findings motivating these behaviors were based on the assumption that children's age was exogenously determined, not addressing bias from selecting into late kindergarten entry.

Positive selection, that which is most commonly associated with redshirting, characterizes parents who view the hold-out year as advantageous; they believe that an additional year at home or in preschool will boost their child's achievement relative to other students, that maturing a year longer will improve children's behavior, or simply that being the oldest in their cohort imbues ongoing positive benefits popularized by Malcom Gladwell's (2008) book, *Outliers*, and perpetuated through social mores (Graue, Kroeger, & Brown, 2003). Redshirted students also typically have substantial socioeconomic advantages, and are more likely to be white and male (Bassok & Reardon, 2013; Schanzenbach & Howard, 2017). On the other hand, parents also delay kindergarten entry when they have concerns about their child's development relative to other four and five year-olds. These negatively selected students are more likely to be diagnosed as having a developmental delay or disability, or are low birthweight (Cook & Kang, 2018; Datar, 2006a; Dhuey, Figlio, Karbownik, & Roth, 2019; Fortner & Jenkins, 2017). These characteristics confound differences in kindergarten entry with differences in other achievement related variables. Such substantial and competing selection factors (and the infeasibility of randomized investigations) explain why there exist no causal impacts of delayed kindergarten entry on human capital outcomes.

A deep and related literature examines the relationships between relative age within cohort, later school achievement, and labor market outcomes. In these studies, the standard identification strategy has been to use quarter of birth or distance from the birthdate cutoff as an instrumental or RD assignment variable to address endogeneity of kindergarten entry age, exploiting state-to state, or between-country differences in birthdate cutoffs for school. [ii] The findings from this literature are mixed, with some studies showing positive impacts on achievement, childhood mental health, reduced grade retention and reduced crime as a teenager (e.g., Bedard & Dhuey, 2006; Black, Devereux, & Salvanes, 2011; Cook & Kang, 2016, 2018; Datar, 2006a; Dee & Sievertsen, 2018; Dhuey et al., 2019; McEwan & Shapiro, 2008), but others find that age-related gains tend to fade out in later grades or by adulthood (Black et al., 2011; Lubotsky & Kaestner, 2016; Robertson, 2011). Other studies identify detrimental impacts of starting school later, with increases in high school dropout rates, reduced educational attainment and years of workforce participation, and thus limited lifetime earnings (Angrist & Krueger, 1991; Black et al., 2011; Deming & Dynarski, 2008; Dobkin & Ferreira, 2010), and some studies still find null effects (Barua & Lang, 2009; Buddelmeyer & Le, 2011; Cascio & Schanzenbach, 2016; Dobkin & Ferreira, 2010). Such paradoxical findings are not unusual. For example, Cook and Kang (2016) find that delayed entrants do better in school and are less likely to become a juvenile delinquent between ages 13-15, but are then more likely to drop out of high school and to commit a crime as an adult. Taken together, the evidence from this literature suggests neither a clear advantage or disadvantage of being amongst the oldest in one's class.

Most recently, Dhuey and colleagues (2019) build upon this body of work using rich, census-level linked birth and school records from Florida to examine difference in outcomes between students born in September (i.e., oldest in cohort) versus August (i.e., youngest) using RD with family fixed effects to robustly control for selection factors. They find a consistent

positive impact of being old for grade of 0.2 SD in averaged grades three through eight test scores, which persisted across a wide range of subgroups (i.e., by maternal education, poverty at birth, race/ethnicity, birth weight, gestational age, and school quality). Heterogeneous effects of the August-September difference did appear for disability status, middle and high school course selections, and in kindergarten readiness, though the latter disappear by the time students take their first exams. Their study shows clear advantages of being the oldest compared with being the youngest in grade.

Still, relative age effects are composed of two slightly different treatments: aging or maturity, and cohort position, or "outlier" effects. The latter is distinct from chronological age in that it places students at the top of the class distribution (i.e., the "relative" advantage of relative age), and that for young children especially can involve clear physical differences in size. The size advantage is what Gladwell (2008) underscores in his study that seem to cause perpetual athletic advantages, but such size differences could certainly translate into educational signals to teachers and administrators, especially in the case of disability or academically gifted designations, or for retention decisions. Therefore, the literature on relative age at entry is only partially informative for understanding the effect of delayed entry on achievement and other human capital outcomes.

In addition to the aging and cohort position advantages, redshirted children may also be exposed to additional high-quality early care and education during the gap year, over and above that of their on-time entering peers, compounding the benefits of maturation and age with skill development. Each of these mechanisms would advantage redshirted students, at least temporarily, upon kindergarten entry. We might expect that the contributions to student outcomes from voluntary redshirting vary across student's ages relative to an enrollment cutoff date. Younger students might primarily benefit from cohort position effects while older students

(who would always have been among the oldest of their grade) might primarily benefit from maturation and hold-out experience effects. A key question is whether these advantages are real (i.e., not due to selection), and whether they persist into later grades beyond kindergarten. By leveraging variation in kindergarten entry from an exogenous state-level policy change and examining the causal mechanism behind any redshirting advantages, our study provides an important empirical contribution to understanding this increasingly common parent behavior and state policy reform.

Another related and well-known literature on the importance of early childhood development shows that human capital formation begins very early, that preschool programs contribute to the cultivation of capital in the present and provide the foundation for future formation, and thus schooling should begin as early as possible (Belfield, Nores, Barnett, & Schweinhart, 2006; Cunha & Heckman, 2007; Duncan & Magnuson, 2013). Indeed, many states, including NC, have expanded publicly-funded educational opportunities for young children in recent years with universal or means-tested pre-kindergarten programs (Gormley & Phillips, 2005; Jenkins, 2014; Pianta & Howes, 2009). Children with developmental disabilities may also benefit from early schooling because they are more likely to be identified and offered early intervention (Frey, 2005; Stipek, 2002). In light of this evidence, redshirting may seem irrational; it is costly in terms of child care and lost parental wages, and discounts the importance of early schooling. Our study eliminates the bias from parental behaviors, rational or not, by leveraging an exogenous policy change using comprehensive administrative educational data, providing a strong causal assessment of this behavior and the mechanisms behind any impacts.

III.   Data

Our study uses statewide administrative data collected by the North Carolina Department

Jenkins & Fortner

of Public Instruction (NCDPI) from the 2008-09 (2009) through 2015-2016 (2016) school years. We focus on student achievement in 3rd grade through 5th grade across two cohorts, where 3rd graders in 2014 represent the cohort who began kindergarten in the 2010-11 school year after NC adjusted the enrollment cutoff to September 1st from October 15th, and includes the students who were forced to delay entry. These files contain student information including the child's sex, race/ethnicity, exact birthdate, and their school's identification number for all students in non-charter public schools.

We use student's kindergarten-year enrollment data to identify which students began kindergarten as scheduled in NC public schools, which allows us to correctly identify students who were subjected to the birthdate cutoff policy change, those who did and did not comply with the policy change, as well as those students who voluntarily redshirted. [iii] Compliance with the policy was strong. Among students with complete data who are included in our 3rd grade reading achievement analyses, we observe 929 students who did not comply with the policy change and started school one year early. These non-compliers represent 7.4 percent of the 12,546 students born in the birthdate window affected by the policy change.[iv]

Our achievement outcome variables are end-of-grade (EOG) test scores in reading and mathematics. We standardize the test scores by grade, subject, and year to allow interpretation of differences as standard deviation (SD) units. We also use two curricular variables, designation as having a disability or as academically gifted coded into dichotomous indicators, as outcome variables. Students are coded as being identified with a disability if their school district recorded an exceptionality that includes behavioral, physical, learning, or cognitive disabilities.

Our data also include standard student-level covariates. We use indicators for male, Asian, Black, Hispanic, multiracial, and American Indian students. A set of variables capture measures of student mobility (e.g., moved schools in prior year or within year). Two variables—

one indicating free lunch eligibility and the other reduced-price lunch eligibility—capture a student's family income level. Two variables describe each student's English learner status in third grade; one indicates whether a student currently receives services for English language learners (ELL), and the other indicates whether a student formerly received ELL services. We include the number of days a student is absent as a measure of school engagement. We created a variable that indicates for each student the number of months younger than the birthdate cutoff to control for chronological age.

Table 1 presents descriptive statistics for forced redshirt students, their cohort peers, and pooled descriptives for the two cohorts of students that comprise our OLS analytic sample for 3rd grade reading achievement. The characteristics of students with respect to gifted identification, gender, ethnicity, mobility, absenteeism, income, and English learner status are similar between the two cohorts. Test scores and identification as having a disability, both outcomes of interest, show some differences between the groups, as does age relative to the enrollment cutoff, which is expected from the policy change.

Data missingness is a concern in any analysis using administrative data. We were able to include 82 percent of possible cases in our 3rd grade analysis sample. The most common reason for missing data was from student in-migration (i.e., from another state or private school) between kindergarten and 3rd grade who we do not observe in the kindergarten enrollment files, precluding our confirmation of whether they were subjected to the school entry cutoff policy change. These students comprise 78 percent of the cases excluded due to missing data. Among students who we observe in our kindergarten roster data and in 3rd grade, we excluded 8,076 records because they were missing other student characteristics. These students represented about 4.5 percent of all records appearing in the kindergarten enrollment files and our 3rd grade records.

We present descriptive comparisons for students included or excluded from our 3rd grade OLS reading models based on missing data in Online Appendix Table 1, which show few systematic differences between the two samples. Students with disabilities are less likely to participate in state testing and students identified as partial year enrollees are also less likely to have testing records in our data. Note that our data only represent students attending public schools in NC, and not private school kindergarten enrollment. However, Dhuey et al. (2019) find that that selection into public schools is unlikely to pose empirical problems.

IV.   Empirical Strategy

Our study uses a policy change in the enrollment cutoff in NC to identify the effect of delaying kindergarten entry on achievement outcomes and to understand the mechanism underlying such policy changes on affected students. Figure 1 presents a diagram of the cohorts and students who were and were not directly impacted by the enrollment cutoff policy change. Shaded in dark grey are the cohorts affected by the policy change; cohorts in light grey are unaffected by the change and are shown to illustrate the shift in cohort age. In the 2008-09 school year, students born between October 16, 2002 and October 15, 2003 were eligible to enroll in kindergarten. For children entering school in the 2009-2010 school year, the state legislature moved the kindergarten birthdate cutoff up by six weeks (Oct 15th to September 1st), requiring that children turn five years of age by September 1st, on approximately the first day of classes. The treatment group comprises the "forced redshirt" students, those born between September 1, 2004 to October 15, 2004 indicated by the black shaded boxes, whose entry into kindergarten was exogenously delayed for one year. Although it was possible for parents to get a waiver for the new policy, there was strong compliance (about 93%), and those who entered

school early (as scheduled before the change) with the 2009 kindergarten cohort are the non-compliers. We assess the extent to which compliance affects our results in Section VI.A.

This change permanently increased the average age of kindergarten cohorts to be approximately 45 days older than prior cohorts, and the transition to the new cutoff would have several impacts on students in the 2009-10 and 2010-11 entry cohorts. Mechanically, this shift of forced redshirts reduced the 2009-10 cohort size. It also meant that the average age of students in the 2009-10 cohort remained younger than the 2010-11 cohort, and displaced the students born on October 16, 2004 from being the oldest cohort members. These features of the policy change and the use of birthdates for treatment assignment are central to our estimation strategy and comparison group design.

A. Comparison Groups and Estimation Methods

The first and perhaps simplest comparison group comprises all the other younger students in the forced redshirt's 2010-2011 entry cohort; these students are the peers of our forced redshirts and experience the same classrooms and testing environment, which include all students shown in the dark grey shaded bins in the 2010-11 row of Figure 1. In Figure 2, we illustrate four additional and more nuanced comparison groups for forced redshirt students, who are shown as group A. The comparison commonly used in the relative age literature is the group of the youngest students in the 2010-2011 cohort, those born between July 15, 2005 and August 31, 2005, shown as group E, to construct something similar to a September to August difference in outcomes as in Dhuey et al. (2019). A third comparison group is the next 45 days-worth of slightly younger students in the forced redshirts' 2010-11 cohort who were not subjected to the policy change (group D); these students are very close to our treatment group in terms of age and in cohort position, born between October 16, 2004 and November 30, 2004. Another comparison is with the 45 days-worth of students born between October 16, 2003 and November 30, 2003

(group B), who comprise the oldest students in the *prior* cohort who were not subjected to the policy change, are very close to forced redshirts in terms of age (at the time of grade 3 testing) and are in the same cohort position (i.e., oldest). Yet another possibility is the 45 days-worth of slightly older students in the prior cohort, those born between July 15, 2004 and August 31, 2004 (group C), that would have been the cohort peers of our treated group in the absence of the birthdate policy change.

Each of the comparison groups have limitations. The other students in the 2010-11 cohort vary widely in age and in cohort position. The "oldest to youngest" comparison a la Dhuey et al. (2019; group E) does not isolate the effect of forced redshirting per se because the youngest in cohort peers differ both in age, cohort position, and in experiencing a hold-out year. The "next 45 days-worth" of younger students (group D) share the same cohort characteristics and experience identical testing instruments, but are not the oldest members of their cohort. The oldest members of the prior cohort (group B) are slightly younger than forced redshirts at the time of testing, are enrolled in a smaller than average cohort, and are tested in a different year. The slightly older students in the prior cohort group who just missed the new birthdate cutoff (group C) are the youngest members of their cohort, are enrolled in a smaller than average cohort, and experience testing and other outcomes a year prior to forced redshirts. Depending on the comparison group selected and estimation strategy we use, treatment effects are collinear with year effects, cohort position effects, and cohort size effects that cannot be controlled. We approached our identification strategy with each of these confounds in mind to explore a number of possible estimates of treatment effects. To best isolate the impact of delayed entry and to triangulate the causal mechanisms behind potential impacts, we use a multiple comparison method approach, described as follows.

1. Forced Redshirt Intent-to-Treat (OLS)

Jenkins & Fortner

We first estimate a simple pooled cross-sectional OLS model of the two contiguous cohorts of interest (the cohort before and the cohort subject to the policy change, shown in dark grey in Figure 1), and our key coefficient of interest is the indicator for forced redshirts, which equals one for the 45 days-worth of students exogenously subjected to delayed entry in the Fall of 2010, controlling for a set of observable characteristics. Because of the policy change, forced redshirt status is as good as randomly assigned, and therefore this straightforward model provides us with an estimate of the intent-to-treat effect (ITT) of delayed kindergarten entry. The specification for our analysis of 3rd grade outcomes is as follows:

1. $Y_{is} = \beta_1 FR_i + X_i \gamma + \theta_s + 2014 + \varepsilon_{is}$

where Y_{is} denotes the outcome (reading and math achievement scores, disability or gifted designations) for or individual i in school s. FR_i is a dichotomous variable equal to 1 if a student was born after the new kindergarten birthdate cutoff but before the old cutoff and was a forced redshirt, and zero otherwise. X_i represents the set of student-level covariates. School fixed effects, θ_s, capture time-invariant school-level factors that influence achievement and curricular decisions; policies for gifted and disability identifications are determined at the district and school levels, and therefore the school fixed effects control for differences in identification practices. School-year fixed effects are captured in the indicator 2014 that accounts for unobserved factors common to students tested in a specific year (i.e., changes to statewide assessments; 4th and 5th grade outcome analyses would substitute 2015 and 2016, respectively). Robust standard errors are clustered at the school level. β_1 is the coefficient of interest and represents the difference in outcomes between forced redshirts and their relatively younger peers in their newly assigned cohort as an ITT effect. For brevity, we conduct our analyses of heterogeneity using this specification only for each of the key subgroups (sex, race (white, black, Hispanic), income) using interactions between each group and FR_i. These estimates are the most

straightforward and provide us with the largest possible sample size from which we can divide

the students into subgroups with enough power to detect heterogeneous effects.

2. Oldest to Youngest in Cohort Comparison (Difference-in-Differences)

Our next comparison group is the youngest students in the cohort (Figure 2, Group E). To

estimate this difference and to exploit the quasi-random assignment to oldest in our treatment

cohort, we construct a difference-in-differences (DID) model for examining 3rd grade outcomes,

modifying equation 1 as follows:

2. $Y_{is} = \beta_1 Old_i + \beta_{Post} 2014 + \delta_{Old*Post} Old_i * 2014 + \beta_2 Middle_i + X_i \gamma + \theta_s + \varepsilon_{is}$

where Old_i is a dichotomous variable indicating students who are the oldest in their cohort (45

days-worth), which in the 2013-14 school year, are the forced redshirts. The coefficient of

interest is δ, which represents the effect of being the oldest as a result of forced redshirting

compared to the reference group, the youngest 45-days worth of students in the 2014 school

year, with those students born in between indicated by $Middle_i$. This difference in comparison

group generates an alternate ITT. All other terms remain the same as above.

3. "Next 45 days" in Cohort Comparison (Regression Discontinuity)

Our next two models begin to investigate the causal mechanisms operating behind

redshirting effects. Here, we restrict the comparison group to the next 45 days of students born

just after the old birthdate cutoff (Oct 16th, 2004) who were not subjected to the policy change

using regression discontinuity (RD) (Figure 2, group D). This comparison removes from our

estimate of interest, FR_i, the effects of aging using birthdate as the assignment variable. We

estimate both the global polynomial OLS and nonparametric local polynomial specifications

using a 45-day bandwidth that corresponds precisely with treatment assignment for forced

redshirts to the left of the October 15th cutoff, both with and without covariates, though our

results are robust to their inclusion. For the local polynomial RD, we use a triangular kernel and

bias-corrected inference procedures developed by Calonico, Cattaneo, and Titiunik (2018; 2014) estimated in STATA with the rdrobust command suite (Calonico, Cattaneo, Farrell, & Titiunik, 2017). Although the consensus in the technical literature supports the local polynomial for RD estimates, we present the OLS version in our main results table so that they can be compared with our other specifications (i.e., including school fixed effects), with local polynomial results presented in Online Appendix Table 4.[v] These estimates represent a local average treatment effect of forced redshirting for students who complied with the policy. To check for smoothness in the covariates across the cutoff, we present descriptive comparisons for students on either side of the Oct 15th cutoff in Online Appendix Table 2. Children on both sides of the cutoff are very similar across the full set of observable characteristics. A McCrary (2008) density test returned a statistically insignificant result indicating no bunching of observations around the cutoff.

4. "Next 45 Days" Difference-in-Discontinuities (RD + DID)

Though our OLS models control for months younger to adjust for cohort position effects and the RD model in (3) restricts comparisons to the top of the cohort, we wanted to more directly assess the extent to which our treatment effect estimates reflect the impact of the redshirt hold-out year (i.e., skill development) or being moved to the top of the age distribution in the cohort because the birthdate policy change caused both treatments. Our final specification builds upon approaches (2) and (3) to better examine these mechanisms. Here, we compare the oldest children to the next oldest 45 days of children using RD (Figure 2, group A to group D), while also making this same RD comparison in the *prior* cohort—for 90 days-worth of children who are the oldest in their cohort and were unaffected by the policy change (Figure 2, group B to children born 12/01-1/15 (not shown)). We then compare these two separate discontinuities across the two cohorts with a difference-in-discontinuities design (Grembi, Nannicini, & Troiano, 2012). This allows us to fully control for cohort position, age, and year effects because

the first difference is made within cohort. Any remaining impact of forced redshirting would then be attributed to children's hold-out year experiences which we cannot observe. This specification builds on (2), the standard global polynomial RD specification:

3. $Y_{is} = \beta_1 Old_i + \beta_{Post} 2014 + \delta_{Old*Post} Old_i * 2014 + \beta_2 (DOB_i - Q) + X_i \gamma + \theta_s + \varepsilon_{is}$

which now controls for the quantitative assignment variable, date of birth (DOB), measured in days and centered at the relevant birthdate cutoff Q, for each cohort—Oct 15, 2004 for the cohort who experienced the policy change, and 45-days later for the untreated prior cohort (November 30)—to create symmetric 45-day bandwidths for each oldest to "next-oldest" RD comparison. The key coefficient is δ, the interaction between the cutoff indicator, Old_i, and the 3rd grade testing year of our forced redshirts, 2014, representing the differential effect (discontinuity) of being the oldest as a result of forced redshirting. All other terms remain the same as above.

B. Changes in Counterfactual Conditions

Because the policy change also directly changed cohort sizes, we test for differences in average class size between our treatment and comparison groups. Shown in the Online Appendix Table 3, we find that there was a slight *increase* in class size for our treated cohort, owing to the reduced cohort size for the prior cohort as a result of the policy change. This means that if smaller class sizes are associated with better outcomes, our treatment effect estimates would be slightly downwardly biased, with forced redshirts experiencing somewhat higher class sizes in fourth and fifth grades. Because changes in class size are downstream from the policy change, we did not include this as a control variable. We did however run our analyses including this as a control for comparison, and our results do not change.

Another key difference between the treatment and various comparison groups may be the counterfactual conditions available during the hold-out year. Although we cannot determine the type and quality of care children received in the year prior to kindergarten, we can confirm that

forced redshirts did not receive additional state resources for prekindergarten programming. NC's pre-k program for disadvantaged children, More at Four (now called NC pre-k) is limited to one year of eligibility. The birthdate cutoff for More at Four was moved in the year prior to the elementary school cutoff to accommodate entering children such that they experienced the program in the year prior to kindergarten entry (Williamson, 2009). Therefore, students in the affected cohort were only able to access one year of the state's high-quality early education program.

V.   Results

We present our main results in Table 2. Results are displayed for each of our four primary specifications by grade, and organized into four panels, A-D, for each of our outcome variables. Reading and math achievement coefficients are in SD units. Gifted and disability identification models were estimated using OLS so coefficients can be interpreted as percentage point changes (logit models yield nearly identical coefficient magnitudes when exponentiated).

A. Forced Redshirt ITT (OLS)

Shown in column (1), we find positive ITT impacts of forced delayed kindergarten entry on student performance in both reading and math assessments. Students who delayed entry scored about 0.06 SD higher on 3rd grade reading and math EOG assessments relative to their non-treated peers, holding constant our modeled student characteristics and school differences. These effects persisted through 5th grade, increasing to about 0.08 SD. For curricular designations, forced redshirts were about one percentage point more likely to be identified as gifted relative to similar students within their cohort in grades 4 and 5. Forced redshirts were about 3.5-4 percentage points less likely to be identified as having a disability relative to their cohort peers, consistent across grades. About 13 percent of students in NC public schools are

identified as having a disability (Table 1); therefore the more than three percentage point difference represents a substantial effect of redshirting on a student's likelihood of identification.

B. Oldest to Youngest in Cohort Comparison (Difference-in-Differences)

Displayed in column (2), these estimates provide a direct comparison to the results of relative age studies. While still consistent across grades, the magnitude of the reading and math achievement impacts of forced redshirting are 0.02-0.05 SD lower than our ITT estimates, ranging between 0.04 and 0.06 SD. These are one-quarter of the size of the August-September birthday difference in achievement found by Dhuey et al. (2019), the most rigorously estimated effect, though a key difference is that we examine the oldest and youngest six-weeks' worth of children, corresponding with the forced redshirt treatment. However, we find very similar impacts on disability identification as Dhuey et al. (2019) of around a 0.02 percentage point decrease, also consistent across grades. We find no effects of being the oldest in cohort on identification as gifted.

C. Causal Mechanisms: "Next 45 days" Comparison Within and Across Cohorts (RD and Difference-in-Discontinuities)

Models 3 and 4 start to "decompose" our ITT findings via process of elimination to narrow in on the causal mechanisms behind the impacts of redshirting. We first tighten the comparison group to better control for chronological age by comparing kids who are almost the exact same age but didn't experience a shock to their school entry (i.e., students born just after the old birthdate cutoff, Oct 16th, 2004, who would always enter kindergarten in 2010). These results reveal a clear story; no effects of forced redshirting on any outcome (column 3). Though some small and marginal impacts on reading achievement appear, they are not consistent across grade. This suggests that the causal mechanisms behind redshirting are not the hold-out year experiences (i.e., skill building), but rather are some combination of chronological age and

cohort position, which are somewhat conflated in this estimate. The local polynomial estimates of this RD comparison, without school fixed effects, are shown in Online Appendix Table 4, along with a comparable OLS specification (no school fixed effects). These estimates are similarly null, and hover around zero.

Our next set of results aim to more comprehensively remove the effect of cohort position by making the same RD comparison, but doing so across cohorts, comparing the forced redshirt difference with the cutoff difference from the cohort unaffected by the birthdate policy change. Displayed in column (4), the results show a consistent null effect for reading, math, and gifted identification. These null coefficients are more precisely estimated than the single cohort RD with the improved power of the additional cohort. Suggestive evidence of a reduction in disability identifications appear, around one percentage point, with the 3rd and 5th grade coefficients reaching marginal significance ($p<.10$), and effects for 4th grade statistically significant at conventional levels.

In summary, our examination of causal mechanisms indicates that the benefits from delayed kindergarten entry are almost exclusively due to age and positional advantages from being at the top of the class.

D. Heterogeneity of ITT Estimates

To test for heterogeneity, we estimated our OLS ITT models using a set of interaction terms for sex, race/ethnicity, and income status. Table 3 presents the treatment effects by subgroup. For each model, we interacted forced redshirt status with student characteristics to test for statistically significant differences in treatment impacts across these student groups. Coefficients reflect the p-value of treatment effects estimated separately for each group and the significant interaction column indicates whether the coefficients for the two groups being compared are statistically different from each other.

Jenkins & Fortner

We found no consistent differences in treatment effects for male and female students. Models comparing estimates by race and ethnicity show larger treatment effects for white and black students relative to Hispanic students on nearly every outcome. Black students differentially benefitted from redshirting in terms of reading achievement compared to white students in grades four and five. White students appear to differentially benefit from redshirting with higher rates of gifted identification in grades four and five relative to both black and Hispanic students; this reveals that the overall small, marginally significant treatment effect estimates for gifted in Table 2 masked substantial heterogeneity by race. In fact, forced redshirt students of Hispanic origin were significantly less likely to be identified as gifted in fourth and fifth grades. Although Hispanic students were significantly less likely to be identified as having a disability as a result of delayed entry, their treatment effect was smaller than that of white and black students (about a 2 percentage point decrease for Hispanic students versus 4 percentage point decrease for white and black students in the likelihood of being identified as having a disability).

Encouragingly, we find that treatment effects are consistently higher for economically disadvantaged students on both academic assessments and identification as having a disability. Impacts on test score outcomes are about twice as high for students eligible for free and reduced-price lunch as those for full-price lunch students (0.10 SD vs. 0.05 SD). Low income students who delayed entry were about 4.5 percentage points less likely to be identified as having a disability compared to on-time entry low-income students; full-price lunch students who delayed entry were only about 2.5 percentage points less likely to be identified as having a disability. We did not find consistent differences in gifted identification across income categories.

VI.   Robustness Analyses

A. Adjustments for Compliance

Jenkins & Fortner

Though we observe high compliance rates, we use two approaches to adjust our estimates

for non-compliers and generate treatment on the treated (TOT) estimates. We estimate a Two-

Stage Least Squares (2SLS) model using the policy change as an instrument for enrollment. This

modifies equation 1 above as:

$$2. \quad Y_{is} = \beta_1 \widehat{FR}_i + MY_i + X_i\gamma + \theta_s + 2014 + \varepsilon_{is}$$

where β_1 now represents the compliance adjusted estimate of forced redshirting, $\widehat{FR}$,

generated from a first stage equation predicting compliance (i.e., kindergarten enrollment in the

2010-11 school year), and all other terms remain the same. A second TOT estimate comes from

analyses which restrict the sample to the 2010-11 kindergarten cohort only, which eliminates all

students in the forced redshirt window who entered kindergarten in the prior school year,

violating the new birthdate cutoff policy. Both sets of compliance adjustment analyses are

presented in Online Appendix Table 5.

For all outcomes, we see the same pattern of effects in the 2SLS TOT estimates in

column 1, with coefficients that are slightly larger (~0.01 SD increases above OLS for reading

and math, 1 percentage point increase for disability identification), given strong compliance with

the policy. We also estimated IV probit models for the dichotomous outcomes to assess the

robustness of the OLS estimates, finding nearly identical coefficient magnitudes. The results in

column 2 that include only the treated cohort and omit all non-compliers are nearly identical to

our main OLS results presented in Table 1, column 1.

B. **Compared With the Kids They Would Have Otherwise Gone to School With in the**
   Prior Cohort

Noted above, an alternative approach compares forced redshirts to students who would

have been their closest-in-age peers in the 2009-10 cohort —the counterfactual condition had the

policy change not occurred—with a regression discontinuity design. This exploits distance in

days from the newly assigned birthdate cutoff, comparing forced redshirted children with their "old" peers on the other side of the cutoff in the prior cohort (those born between July 15, 2004 and August 31, 2004; Figure 2, group C). However, recall that this comparison conflates changes in entry year from cohort position, whereby forced redshirts became the oldest in their cohort but their counterfactual peers remained amongst the youngest. For this reason we prefer the set of within-cohort estimates presented in our main results, comparing students to their newly-assigned but relatively younger peers, because we are able to separate the impacts of delayed entry from cohort position and year effects. Even so, this specification does involve an important counterfactual condition. Results are presented in Online Appendix Table 6; the pattern of these RD models are very similar to our OLS estimates, but are consistently two to three times larger for each of the outcomes, comprising changes from a 1-year delayed entry, opposite cohort position, and year effects. Nonetheless, they indicate a strong impact of the combination of such changes on student outcomes.

C. Attrition

Among treated cases, about 87 percent of students appearing in our 3rd grade reading models are also included in our 4th and 5th grade models. A similar percentage, 88 percent, are included in all three years of analysis in the untreated group. We ran models separately for the students who remain in our analysis sample in 3rd through 5th grades, and results were the same as those presented.

VII. Discussion

Our study provides the first causal estimates of the effects of redshirting on achievement outcomes in elementary school. This common practice influences not only the child who delays entry, but also the parents who must provide an additional year of care, the public schools in

Jenkins & Fortner

their efforts to teach kindergarten students with a wider range of ages and pre-school

experiences, and other classmates who did not delay kindergarten entry, with the potential for

equity enhancing or exacerbating effects.

Exploiting a change in the birthdate cutoff policy affecting children entering school in the

fall of 2010, we found, on average, small, positive treatment effects for math and reading

achievement for forced redshirt students, and a 25% reduction in disability identification in 3rd-

5th grades. These effects persist through elementary school, with no evidence of decay. However,

delaying kindergarten entry by one year had no main effect on a child's identification as

academically gifted.

A unique feature of our study was the use of multiple comparison groups to better

understand the causal mechanisms behind redshirting. Here, we found that the overall impacts of

redshirting we detected were not due to hold-out year experiences (i.e., skill accumulation) and

were likely from advantages due to shifting to the oldest position in the cohort and aging effects.

This suggests that redshirting benefits stem from an "outlier" (Gladwell, 2008) effect, whereby

your relative age, size, and perhaps perceived maturity beget comparative advantages versus

your younger cohort peers. Although we consider these analyses to be more exploratory in terms

of identification of mechanisms, it seems important to more precisely communicate to parents

and school districts what may happen when students don't comply with birthdate cutoff policies;

there are some small benefits, but they are entirely positional benefits, meaning that the way

children's days are filled in the hold-out year (and the related expenses from these experiences)

are not likely meaningfully changing their achievement and curricular outcomes. It is difficult to

provide a valuation of achievement outcomes in this context. However, the child care costs or

foregone parental earnings, and the reduced lifetime earnings of the redshirted child from

intentionally delaying a year seem quite large relative to the small effect sizes we find here.

Our heterogeneity analyses revealed some surprising results. Delayed entry differentially benefitted low income students, suggesting some one-time, equity enhancing impacts of birthdate cutoff changes and confirming the exploratory analyses conducted by Dhuey and colleagues (2019). However, forced redshirting did not benefit Hispanic students, primarily benefitted white students, and thus further widened the gap between white and Hispanic achievement. These findings are concerning, especially considering that white families are more likely to redshirt their children, possibly indicating that this practice may have exacerbated gaps in achievement across the country as voluntary redshirting became more widespread. Reducing racial-ethnic and income achievement gaps are both priorities for educational policy. In the case of shifting birthdate cutoffs, such a policy change is not likely to be a pareto improvement or allow states to meaningfully address these important perennial issues in public education.

Although the boosts to reading and math scores that we find as a result of forced redshirting are clearly positive, our finding of reduced risks of identification as disabled may be more difficult to interpret. If reduced identifications are from lower rates of disabilities or delays amongst this group, that would certainly be positive, especially considering the cost of disability services and accommodations, and the stigma associated with identification as having a disability. However, if forced redshirts were in need of services but were less likely to be identified because of their relative age, this could be worrisome, especially considering the long-run benefits of effective early intervention for students with disabilities (Connor, Alberto, Compton, & O'Connor, 2014; National Center for Special Education Research, 2015). Therefore, it is difficult to unequivocally say that reduced rates of identification would result in improved student outcomes. Still, we find overall benefits to math and reading achievement, perhaps reducing concerns of under-identification due to cohort positional effects.

It is important to note that while we leverage a natural experiment to identify a plausibly causal estimate of kindergarten redshirting, the students in our intent-to-treat group all have birthdates within six weeks of the enrollment cutoff. In practice, we would expect that a voluntary decision to redshirt would provide benefits from maturation (one year) plus any cohort shift position effects. For students who will already be in an advantaged cohort position (i.e., early fall birth dates), there would be little to gain from a cohort position mechanism by voluntarily redshirting. Although we cannot precisely disentangle the effects of aging from the effects of cohort position in our study (they are confounded by the treatment), our results suggest that voluntary redshirting may not be beneficial for children who do not receive a positional advantage from voluntary redshirting. Prior studies consistently show that students who will be the youngest members of their kindergarten cohorts are those most likely to be redshirted, perhaps indicating that parents whose children have the most to gain from delaying are those most likely to do so.

The 12-month cohort arrangement of public schooling appears to systematically advantage students who will be among the oldest in their cohort based on the school's entry cutoff, regardless of where in the year this cutoff is placed. An evaluation of this arrangement seems appropriate given the equity implications which provide differential benefits to some students at the expense of others on an arbitrary characteristic of month and day of birth. Policymakers may wish to consider alternative arrangements for schooling which address cohort position advantages.

In sum, our evidence suggests that shifting the kindergarten birthdate cutoff is not, on balance, a strategic or precise policy instrument to improve public education outcomes. At the individual level, families and their children enjoy small improvements in achievement, but these improvements are not the result of investment activities during the hold-out year; rather they are

Jenkins & Fortner

attributed to positional advantages. Is the valuation of the academic benefits of delayed entry

higher than the costs of the hold-out year and the public costs of increased racial-ethnic

achievement gaps? Future research can provide a more precise estimate of this calculation, but

we find this unlikely.

---

[i] In addition, delaying public school enrollment eligibility can have a negative effect on the maternal labor supply, especially for single mothers (Gelbach, 2002). In response to these concerns, states such as California have created transitional kindergarten programs, available for children born in the window between the old and new birthdate cutoffs. However, no published studies have evaluated the effect of these programs. Estimates of the impact of kindergarten birthdate policy changes on student academic outcomes would be a timely contribution to the research and policy literature.

[ii] Although quarter of birth instruments have been widely criticized (Barua & Lang, 2009; Bound, Jaeger, & Baker, 1995), recent work also suggests that the timing of birth is correlated with maternal characteristics (Buckles & Hungerman, 2013). The studies noted here all test for such possibility, but find that covariates are smooth across the birthdate cutoff. Furthermore, in many studies school starting age cannot be disentangled from age-at-test and time-in-school effects.

[iii] We use enrollment comparisons of adjacent kindergarten year data to identify students who were voluntarily redshirted, excluding students who were retained in kindergarten.

[iv] The non-compliance rate based on date of birth only (regardless of complete data) was 7.3 percent. Online Appendix Table 1 provides descriptive statistics that compare students included in our 3rd grade reading models with those excluded due to missing data, revealing no concerning systematic differences in the samples.

[v] We also prefer these estimates because it allows us to make within school comparisons, which are particularly important for curricular designation outcomes.

References

Angrist, J. D., & Krueger, A. B. (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics, 106*(4), 979-1014. doi:10.2307/2937954

Barua, R., & Lang, K. (2009). School entry, educational attainment and quarter of birth: A cautionary tale of LATE. *National Bureau of Economic Research Working Paper Series, No. 15236.*

Bassok, D., & Reardon, S. F. (2013). Academic "Redshirting" in Kindergarten: Prevalence, Patterns, and Implications. *Educational Evaluation and Policy Analysis*. doi:10.3102/0162373713482764

Bedard, K., & Dhuey, E. (2006). The persistence of early childhood maturity: International evidence of long-run age effects. *The Quarterly Journal of Economics, 121*(4), 1437-1472. doi:10.1093/qje/121.4.1437

Belfield, C. R., Nores, M., Barnett, W. S., & Schweinhart, L. (2006). The High/Scope Perry Preschool Program. *Journal of Human Resources, XLI*(1), 162-190. doi:10.3368/jhr.XLI.1.162

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2011). Too young to leave the nest? The effects of school starting age. *The review of economics and statistics, 93*(2), 455-467.

Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association, 90*(430), 443-450.

Buckles, K. S., & Hungerman, D. M. (2013). Season of birth and later outcomes: Old questions, new answers. *Review of Economics and Statistics, 95*(3), 711-724.

Buddelmeyer, H., & Le, T. (2011). *Effects of age at entry to Year 1 on later schooling outcomes: Evidence from Australia*. Melbourne Institute of Applied Economic and Social Research. The University of Melbourne. Melbourne.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal, 17*(2), 372-404.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2018). Regression discontinuity designs using covariates. *Review of Economics and Statistics*(0).

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression‐discontinuity designs. *Econometrica, 82*(6), 2295-2326.

Cascio, E. U., & Schanzenbach, D. W. (2016). First in the Class? Age and the Education Production Function. *Education Finance and Policy, 11*(3), 225-250. Retrieved from http://www.nber.org/papers/w13663

Connor, C., Alberto, P. A., Compton, D. L., & O'Connor, R. E. (2014). *Improving reading outcomes for students with or at risk for reading disabilities: A synthesis of the contributions from the Institute for Education Sciences Research Centers*. (NCSER 2014-3000). Washington, D.C.

Cook, P. J., & Kang, S. (2016). Birthdays, Schooling, and Crime: Regression-Discontinuity Analysis of School Performance, Delinquency, Dropout, and Crime Initiation. *American Economic Journal: Applied Economics, 8*(1), 33-57. doi:doi: 10.1257/app.20140323

Cook, P. J., & Kang, S. (2018). The School-Entry-Age Rule Affects Redshirting Patterns and Resulting Disparities in Achievement. *National Bureau of Economic Research Working Paper Series, No. 24492*. doi:10.3386/w24492

Cunha, F., & Heckman, J. J. (2007). The Technology of Skill Formation. *American Economic Review, 97*(2).

Datar, A. (2006a). Does delaying kindergarten entrance give children a head start? *Economics of Education Review, 25*(1), 43-62. doi:10.1016/j.econedurev.2004.10.004

Datar, A. (2006b). The impact of kindergarten entrance age policies on the childcare needs of families. *Journal of Policy Analysis and Management, 25*(1), 129-153. doi:10.1002/pam.20159

Dee, T. S., & Sievertsen, H. H. (2018). The gift of time? School starting age and mental health. *Health Economics, 27*(5), 781-802. doi:doi:10.1002/hec.3638

Deming, D., & Dynarski, S. (2008). The lengthening of childhood. *Journal of Economic Perspectives, 22*(3), 71-92.

Dhuey, E., Figlio, D., Karbownik, K., & Roth, J. (2019). School Starting Age and Cognitive Development. *Journal of Policy Analysis and Management, 38*(3), 538-578. doi:10.1002/pam.22135

Dobkin, C., & Ferreira, F. (2010). Do school entry laws affect educational attainment and labor market outcomes? *Economics of Education Review, 29*(1), 40-54. doi:10.1016/j.econedurev.2009.04.003

Duncan, G. J., & Magnuson, K. (2013). Investing in preschool programs. *The Journal of Economic Perspectives, 27*(2), 109-132. doi:10.1257/jep.27.2.109

Education Commission of the States. (2014). Kindergarten Entrance Age. Retrieved from http://ecs.force.com/mbdata/mbquestRT?rep=Kq1402.

Elder, T. E., & Lubotsky, D. H. (2009). Kindergarten entrance age and children's achievement. *Journal of Human Resources, 44*(3), 641-683. Retrieved from http://jhr.uwpress.org/content/44/3/641.abstract

Fortner, C. K., & Jenkins, J. M. (2017). Kindergarten redshirting: Motivations and spillovers using census-level data. *Early Childhood Research Quarterly, 38*, 44-56.

Frey, N. (2005). Retention, Social Promotion, and Academic Redshirting. *Remedial and Special Education, 26*(6), 332-346. doi:10.1177/07419325050260060401

Fryer, R. G., & Levitt, S. D. (2006). The Black-White Test Score Gap Through Third Grade. *American law and economics review, 8*(2), 249-281. doi:10.1093/aler/ahl003 pmid:

Gelbach, J. B. (2002). Public schooling for young children and maternal labor supply. *American Economic Review*, 307-322.

Gladwell, M. (2008). *Outliers: The story of success*. UK: Hachette.

Gormley, W. T., & Phillips, D. (2005). The Effects of Universal Pre-K in Oklahoma: Research Highlights and Policy Implications. *Policy Studies Journal, 33*(1), 65-82.

Graue, E., Kroeger, J., & Brown, C. (2003). The gift of time: Enactments of developmental thought in early childhood practice. *Early Childhood Research & Practice*(Spring).

Grembi, V., Nannicini, T., & Troiano, U. (2012). *Policy responses to fiscal restraints: A difference-in-discontinuities design* (3999). Retrieved from http://ssrn.com/abstract=1852523

Jenkins, J. M. (2014). Early childhood development as economic development: Considerations for state-level policy innovation and experimentation. *Economic Development Quarterly, 28*(1).

Lubotsky, D., & Kaestner, R. (2016). 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, 53*, 194-206. doi:http://dx.doi.org/10.1016/j.econedurev.2016.04.001

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics, 142*(2), 698-714. Retrieved from http://www.sciencedirect.com/science/article/pii/S0304407607001133

McEwan, P. J., & Shapiro, J. S. (2008). The Benefits of Delayed Primary School Enrollment: Discontinuity Estimates Using Exact Birth Dates. *Journal of Human Resources, 43*(1), 1-29. doi:10.3368/jhr.43.1.1

Meisels, S. J. (1999). Assessing readiness. In R. Pianta & M. J. Cox (Eds.), *The transition to kindergarten* (pp. 39-66). Baltimore: Paul H. Brookes.

Morrison, F. J., Alberts, D. M., & Griffith, E. M. (1997). Nature-nurture in the classroom: Entrance age, school readiness, and learning in children. *Developmental Psychology, 33*(2), 254-262. doi:10.1037/0012-1649.33.2.254

National Center for Special Education Research. (2015). *Investment in Reading Research from Kindergarten through High School*. Institute for Education Sciences, Department of Education.

Paul, P. (2010, August 20). The littlest redshirts sit out kindergarten. *New York Times*.

Pianta, R., & Howes, C. (Eds.). (2009). *The promise of pre-k*. Baltimore, MD: Paul H. Brookes Publishing Co.

Robertson, E. (2011). The effects of quarter of birth on academic outcomes at the elementary school level. *Economics of Education Review, 30*(2), 300-311. doi:http://dx.doi.org/10.1016/j.econedurev.2010.10.005

Schanzenbach, D. W., & Howard, S. L. (2017). Is Your Child Ready for Kindergarten? *Education Next, 17*(3), 18-24.

Stipek, D. (2002). At what age should children enter kindergarten? : A question for policy makers and parents. *SRCD Social Policy Report, XVI*(2). Retrieved from http://www.srcd.org/Documents/Publications/SPR/sprv16n2.pdf

Williamson, S. N. (2009, February 27, 2009). Is your child old enough for kindergarten? *Hickory Record*.

Entry Year

Birthdate ranges for kindergarten cohorts

Entry Year										
2008-09			10/16/2002							10/15/2003
2009-10			10/16/2003					8/30/2004	9/1/2004	10/15/2004
2010-11			10/16/2004					8/30/2005		
2011-12	9/1/2005							8/30/2006		

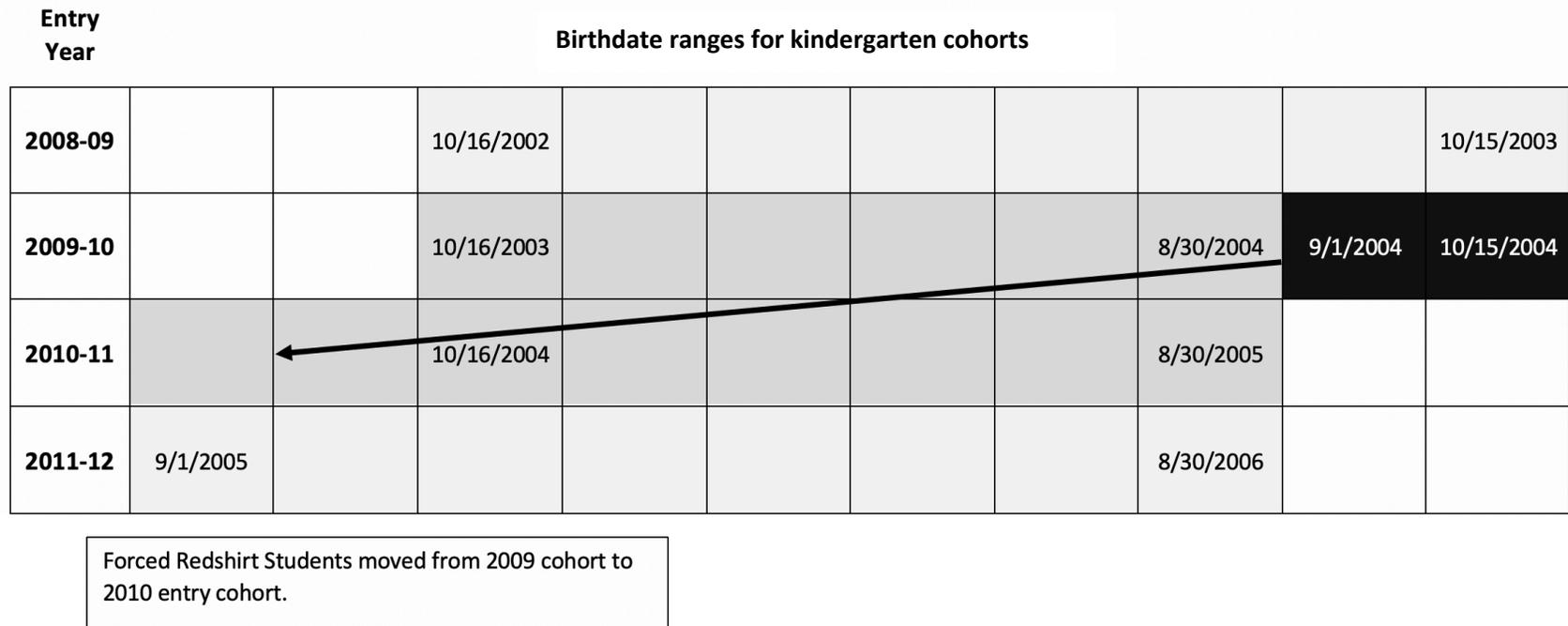Forced Redshirt Students moved from 2009 cohort to 2010 entry cohort.

Figure 1: Kindergarten cohort changes over time due to the NC policy change

Caption: This diagram includes the cohorts and students who were and were not directly impacted by the enrollment cutoff policy change in 2009. Cohorts shaded in dark grey were affected by the policy change; cohorts in light grey are unaffected by the change. Forced redshirt students, those born between September 1, 2004 to October 15, 2004, are indicated by the black shaded boxes.
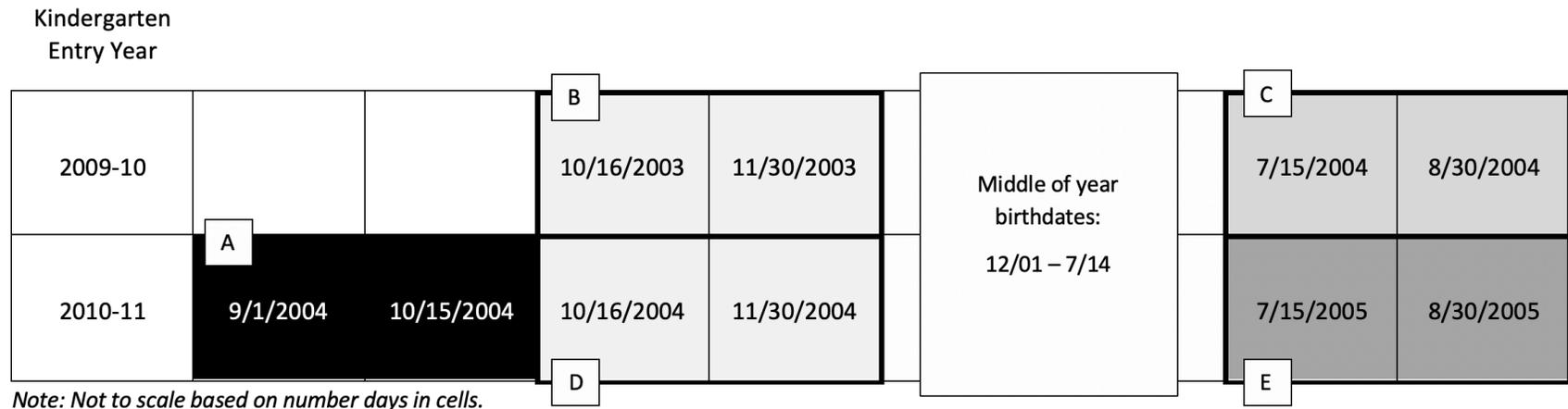
Kindergarten
Entry Year

		B				C	
2009-10		10/16/2003	11/30/2003	Middle of year birthdates:		7/15/2004	8/30/2004
2010-11	A 9/1/2004 10/15/2004	10/16/2004	11/30/2004	12/01 – 7/14		7/15/2005	8/30/2005
	D					E	

Note: Not to scale based on number days in cells.

Figure 2: Comparison group possibilities

Table 1: Descriptive Statistics

	Forced Redshirts	Within-cohort comparison	Diff	Full OLS Analytic sample
Reading Score (std)	0.17	-0.03	0.21	-0.01
Math Score (std)	0.17	-0.01	0.19	0.01
Gifted Identification	0.06	0.06	0.01	0.08
Disability Identification	0.09	0.13	-0.03	0.12
Forced Redshirt (Intent-to-treat)	1.00	0.00	1.00	0.07
Months Younger Than Age Cutoff	0.74	6.85	-6.10	5.92
Months Younger Than Age Cutoff Sq.	0.74	56.58	-55.84	46.97
Male	0.51	0.50	0.00	0.51
Black	0.24	0.25	-0.02	0.24
Hispanic	0.16	0.16	0.00	0.16
Asian	0.03	0.03	0.00	0.03
Am. Indian	0.01	0.01	0.00	0.01
Multiracial	0.04	0.04	0.00	0.04
Moved Within Year	0.08	0.07	0.01	0.06
Moved Prior Year	0.00	0.00	0.00	0.00
Days Absent	4.97	4.96	0.01	5.80
Free Lunch Eligible	0.50	0.52	-0.02	0.51
Reduced Price Lunch Eligible	0.06	0.06	0.00	0.06
English Language Learner	0.09	0.10	-0.02	0.10
English Language Learner (Former)	0.01	0.01	0.00	0.01
Redshirt (Voluntary)	0.00	0.02	-0.02	0.02
N	11,617	78,471		172,826

Note: Descriptive statistics for observations included in our OLS 3rd grade reading model.

Table 2: Estimated Effects of Forced Redshirting on Outcomes in 3rd Through 5th Grade Using Multiple Comparisons

Grade level	(1) Forced RS compared with all other students in cohort (OLS ITT)	(2) Oldest compared with youngest in cohort (Diff-in-Diff)	(3) Forced RS compared with next 45 days in cohort (RD)	(4) Diff-in-Discontinuities between oldest and next 45 days in cohort (RD + Diff-in-Diff)
A. Reading achievement				
3	0.057**	0.045**	0.032	0.011
	(0.010)	(0.012)	(0.024)	(0.018)
4	0.081**	0.056**	0.031	0.026
	(0.010)	(0.012)	(0.024)	(0.018)
5	0.083**	0.061**	0.006	0.019
	(0.009)	(0.012)	(0.024)	(0.017)
N	172826	172837	21678	41187
B. Math achievement				
3	0.060**	0.049**	0.011	0.011
	(0.009)	(0.012)	(0.024)	(0.017)
4	0.095**	0.061**	0.020	0.017
	(0.010)	(0.012)	(0.024)	(0.017)
5	0.086**	0.042**	0.004	0.006
	(0.009)	(0.012)	(0.024)	(0.017)
N	174006	174017	21811	41388
C. Gifted Identification				
3	0.004	-0.000	-0.012+	-0.002
	(0.003)	(0.004)	(0.006)	(0.005)
4	0.011**	0.007	-0.013	-0.003
	(0.004)	(0.005)	(0.009)	(0.007)
5	0.012**	0.006	-0.008	-0.005
	(0.004)	(0.005)	(0.010)	(0.008)
N	176102	176113	22139	41743
D. Disability Identification				
3	-0.033**	-0.020**	0.009	-0.011+
	(0.003)	(0.005)	(0.009)	(0.006)
4	-0.041**	-0.023**	0.005	-0.013*
	(0.004)	(0.005)	(0.009)	(0.007)
5	-0.038**	-0.017**	0.006	-0.011+
	(0.004)	(0.005)	(0.010)	(0.007)
N	176102	176113	22139	41743

Notes: Observation counts are from 3rd grade models. Reading and math achievement are in standard deviation units, and gifted and disability identification are in percentage point units. All models include school fixed effects, and the set of covariates described in the text and shown in Table 1. Robust standard errors (in parentheses) are clustered at the school level. ** $p < 0.01$, * $p < 0.05$, + $p < 0.10$

Jenkins & Fortner

Table 3. Treatment Effects by Subgroup

Grade level	Sex			Race					Income		
	Female	Male	Statistically Significant Interaction	White	Black	Statistically Significant Interaction	Hispanic	Statistically Significant Interaction	Full Price Lunch	Free/Red. Price Lunch	Statistically Significant Interaction
A. Reading achievement											
3	0.049** (0.012)	0.065** (0.013)	-	0.060** (0.012)	0.067** (0.018)	-	0.025 (0.023)	-	0.039** (0.012)	0.074** (0.013)	*
4	0.085** (0.012)	0.052** (0.013)	*	0.069** (0.012)	0.115** (0.017)	*	0.001 (0.020)	**	0.038** (0.012)	0.101** (0.013)	**
5	0.085** (0.012)	0.076** (0.013)	-	0.074** (0.012)	0.141** (0.016)	**	0.018 (0.018)	**	0.052** (0.012)	0.112** (0.013)	**
N	172826			172826			172826		172826		
B. Math achievement											
3	0.064** (0.012)	0.053** (0.012)	-	0.070** (0.011)	0.058** (0.018)	-	0.011 (0.022)	*	0.041** (0.013)	0.074** (0.012)	*
4	0.099** (0.012)	0.064** (0.013)	*	0.087** (0.012)	0.105** (0.017)	-	0.028 (0.020)	**	0.057** (0.012)	0.107** (0.013)	**
5	0.077** (0.012)	0.080** (0.013)	-	0.087** (0.012)	0.111** (0.017)	-	0.002 (0.019)	**	0.050** (0.012)	0.109** (0.013)	**
N	174006			174006			174006		174006		
C. Gifted Identification											
3	-0.000 (0.004)	0.008* (0.004)	+	-0.000 (0.004)	0.015** (0.004)	**	0.001 (0.005)	-	-0.010* (0.005)	0.016** (0.003)	**
4	0.007 (0.005)	0.005 (0.005)	-	0.015** (0.005)	0.001 (0.005)	*	-0.016* (0.006)	**	0.005 (0.006)	0.008+ (0.004)	-
5	0.008 (0.005)	0.009 (0.005)	-	0.018** (0.006)	0.001 (0.006)	*	-0.013+ (0.007)	**	0.011+ (0.006)	0.006 (0.005)	-
N	176102			176102			176102		176102		
D. Disability Identification											
3	-0.028** (0.004)	-0.040** (0.005)	*	-0.036** (0.004)	-0.039** (0.006)	-	-0.017* (0.007)	*	-0.020** (0.004)	-0.045** (0.004)	**
4	-0.032** (0.004)	-0.041** (0.005)	-	-0.040** (0.004)	-0.038** (0.006)	-	-0.022** (0.007)	*	-0.027** (0.004)	-0.046** (0.005)	**
5	-0.034** (0.004)	-0.044** (0.005)	-	-0.041** (0.004)	-0.047** (0.006)	-	-0.022** (0.007)	**	-0.027** (0.004)	-0.053** (0.005)	**
N	176102			176102			176102		176102		

Notes: Observation counts from pooled 3rd grade models that include interaction terms between forced redshirting and each student characteristic. Black and Hispanic students are compared with white students separately. Reading and math achievement are in standard deviation units, and gifted and disability identification are in percentage point units. All models include school fixed effects, and the set of covariates described in the text and shown in Table 1. Robust standard errors (in parentheses) are clustered at the school level. ** $p<0.01$, * $p<0.05$, + $p<0.10$