The Impact of School Entry Age on Student Achievement: Evidence from Nebraska

Ishita Ahmed*

October 27, 2023

Abstract

Using administrative data with exact date of birth from the Nebraska Department of Education, I assess the impact of waiting an additional year to start kindergarten on students' educational outcomes and find a positive significant impact of waiting on test scores. The positive impact of the fuzzy regression discontinuity design diminishes over time, and the diminishing effect is more pronounced for children from disadvantaged households. This suggests that the decision to delay kindergarten may worsen the socioeconomic achievement gap. However, using exogenous change in the kindergarten entry policy in Nebraska, I could not find any impact of moving the kindergarten cutoff earlier on the achievement gap. A reduction in the practice of redshirting might drive that result.

JEL Codes: I24, I28, J15

*Ahmed:Department of Economics, College of Business, University of Nebraska-Lincoln ishita.ahmed@huskers.unl.edu

1 Introduction

School readiness holds considerable importance in determining academic achievement. Given the absence of a clear-cut measure for defining school readiness, the age at which a child enters school is commonly used as a proxy of their readiness. A simple strategy to improve achievement score is to raise the average school entry age (Stipek, 2002). This can be achieved by moving the kindergarten cutoff earlier or advising parents to hold back their children for a year from attending kindergarten (Lincove & Painter, 2006). It is believed that older students exhibit greater maturity and gain an academic advantage, which contributes to their improved performance (Uphoff & Gilmore, 1985). Numerous studies have investigated the effects of school entry age on test scores. While analyzing the influence of entry age provides valuable insights, the emphasis is mainly on the average impact. However, this standard approach might overlook certain factors that carry significance for policymaking. For example, understanding how the impact of entry age varies across different academic achievement levels is vital. The decision to delay entering kindergarten may affect high-scoring students differently than those with lower scores. Additionally, this impact may differ across demographic characteristics. Studies indicate that the entry age effects can vary for a child from a wealthy household compared to one from a poorer background (Elder & Lubotsky, 2009; Datar, 2006). Elder & Lubotsky (2009) attributes the difference in impact to access to quality daycare, suggesting that a child from a wealthier family might have attended a good daycare before starting kindergarten, thereby accumulating some human capital. Additionally, many studies have demonstrated the diminishing effect of entry age as students progress through school, but little attention has been given to whether this fading pattern exhibits heterogeneity based on their background. Considering all these factors, it is crucial to address some additional questions that have not been explored yet: For whom does entry age matter? Does the impact show a fading pattern, and does the diminishing rate exhibit uniformity across different demographic groups?

This paper utilizes longitudinal data at the state-level on nine cohorts from Nebraska to estimate the impact of entry age on achievement scores from kindergarten to middle school level. The cohorts began kindergarten between the academic years 2007-2008 and 2015-2016. I specifically use fuzzy regression discontinuity framework which leverages exogenous variation in school starting age imposed by the state law to estimate the impact of waiting an additional year to start school on test scores from kindergarten through 8th grade. My analysis is based on the student-level administrative data from the Nebraska Department of Education with students' precise date of birth. I find waiting an additional year to start kindergarten increases test scores by 0.3 standard deviation. I also run IV quantile regression to investigate the impact of entry age throughout different levels of achievement score. I find the impact to be particularly pronounced among the top performers in the class. In addition to that I check if there is any evidence of heterogeneity in impact for each grade by demographic characteristics, precisely if the impact varies by gender, socioeconomic status, or race. The advantage of waiting is more evident in advantaged groups, aligning with the findings of Elder & Lubotsky (2009). While the achievement gap between older and younger students gradually diminishes, the rate of reduction is notably faster for children from disadvantaged backgrounds. The distinct magnitude and pattern of this reduction might contribute to an acceleration of the socioeconomic and racial achievement gap. Considering the change in kindergarten entry policy in Nebraska, I explore how moving the kindergarten cutoff earlier, in the calendar year, affects the achievement gap between advantaged and disadvantaged groups. Surprisingly, I find that the policy reform does not have a substantial impact on the achievement gap. This result could be driven by a growing gap resulting from waiting and a reduction in the practice of redshirting. Importantly, my results remain consistent across various specifications.

My paper is related to three different streams of literature. First, I add to the literature on kindergarten eligibility for different academic achievements. Most studies have studied the association between school entry ages and various educational outcomes. Though usually a higher entry age gives a comparative advantage in school, ranging from higher test scores (Datar, 2006;Attar & Cohen-Zada, 2018; Elder & Lubotsky,2009; Cook & Kang, 2016; Fletcher & Kim, 2016) to a lower probability of grade retention (Cook & Kang, 2016). A higher entry age also leads to higher probability of High-School dropout (Cook & Kang, 2016) and lower human capital accumulation as entering school late causes to reach mandatory school attendance age early. There are studies that investigate the impact on other outcomes, like adult wages (Bedard & Dhuey, 2007), juvenile crimes (Depew & Eren, 2016), and felony offense (Cook & Kang, 2016). All these studies have estimated the mean effect of school entry age on outcomes. Unfortunately, the conventional approach does not take into account the heterogeneity in impact across score distribution. To overcome that, I incorporate quantile approach in the school readiness literature, which has not been studied before. I find waiting to enter kindergarten to increase the achievement score consistent with previous works on birthday effects and academic performances. This quantile approach additionally enables me to observe considerable heterogeneity throughout the distribution. The impact of waiting is larger for those who are at the top of the score distribution.

This paper contributes to the school readiness literature by comparing the progress of various subgroups based on demographic characteristics throughout different grades. Some work finds that the impact of entry age does not persist for long, and the achievement gap between older and younger children starts to fade away (Lincove & Painter, 2006; Elder & Lubotsky, 2009). Considering the gap in the magnitude of impact among various groups, I investigate the pattern of fading away of entry-age effects across different demographic groups as they progress through school. My findings reveal that the age effect persists longer among children from high-income families than other groups, exacerbating the achievement gap. This happens due to inadequate support for developing children from low-income households. Thus, incorporating these distinct trajectories, we gain insights into the underlying mechanisms contributing to this phenomenon.

This paper also contributes to the literature on the achievement gap. Previous studies on the achievement gap have mainly explored the effects of early childhood influences and other factors such as school standards, cultural norms, and childhood environment on the achievement gap (Autor et al., 2016; Nollenberger et al., 2016; Chetty et al., 2016). Some studies have also investigated the impact of entry age on the achievement gap. For example, Oshima & Domaleski (2006) identified the existence of an achievement gap up to grade 5. My work is closely connected to the research of Lenard & Peña (2018). They used an instrumental variable strategy based on policy reform in North Carolina and found an increase in the minority-majority achievement gap (based on race) in waiting. However, Lenard & Peña (2018) focused on only one school district and the test scores of grade three. I contribute to this literature by specifically focusing on the socioeconomic achievement gap and incorporating data from multiple grades for analysis.

The rest of the paper proceeds as follows. In the next section, I provide details of the institutional background, data, and descriptive statistics. Section III describes the intuitions behind the research design. Section IV explores the empirical strategy. Section V presents my results, followed by the impact of policy reform in section VI. Section VII discusses robustness checks and sensitivity analyses. In the end, I conclude in section VIII.

2 Settings and Data

2.1 Admission Process

Before the 2012-13 school year, the cutoff date for kindergarten enrollment in Nebraska was October 15. This means that children had to turn five years old on or before October 15 in order to be eligible for kindergarten. However, the provision of early entrance has always been there in the statute of Nebraska. Before the reform, children born between October 16 and February 1 were allowed to start school conditional on some requirements. [1] However, starting from the 2012-13 school year, the cutoff date was changed to July 31, allowing children to enter kindergarten if they reached the age of five on or before that date. Children born on Aug 1 or later, will be deferred and will start kindergarten almost at the age of six. However, as mentioned before there are opportunities for early entry. Children who turn five between two cutoffs, precisely, August 1 and October 15, might also be considered eligible if they demonstrate satisfactory performance on a school assessment. This assessment evaluates whether a child possesses the necessary skills to handle kindergarten-level tasks (Neb. Rev. Stat. § 79-214). Regulations on determining maturity depends a lot on parents' discretion. Parents might accelerate kindergarten entry if they feel the child is skilled enough to start early. In contrast, if parents suspect any developmental delay, they might hold back their children even if they are eligible to enter. Particularly, the reason of redshirting or greenshirting is not possible to determine (Fertig & Kluve, 2005). This leads to variation in enrollment practices.

[1]"A board can admit a student even if he/she is born after the cutoff if parent or guardian request and provides an affidavit stating that (a) the child attended kindergarten in another jurisdiction in the current school year, (b) the family anticipates relocation to another jurisdiction that would allow admission within the current year, or (c) the child has demonstrated through recognized assessment procedures approved by the board that he or she is capable of carrying the work of kindergarten or the beginner grade" (Neb. Rev. Stat. § 79-214)

2.2 Sample

The current study uses student-level administrative data collected by the Nebraska Department of Education (NDE). The sample spans data from 2007-2008 to 2018-2019. Due to the COVID, I restrict my sample up to 2018-2019. However, it includes all first-time kindergarten entrants from 2007-2008 to 2015-2016 who enrolled in the public schools of Nebraska. The data is limited to only students from public schools in Nebraska, so they are supposed to follow the state regulation of kindergarten entry. The kindergarten entry cohort ranges from around twenty-one to twenty-two thousand each year. I excluded 14,126 observations as they had incongruous birth dates across different grades. As it is longitudinal data, I can follow each kindergarten entrant for a few years until grade eight[2]. One limitation of the data is that it excludes children studying at private schools. Furthermore, I cannot observe if any student moves out of the state. I restrict the samples' entry age from 57 to 79 months during August 1 of their kindergarten entry year as in Nebraska compulsory school age is six by January 1 (Neb. Rev. Ann. § 79-201). I also drop those students I cannot observe in kindergarten to estimate their age at entry precisely. Of my sample of first-time kindergarten entrants, 48 percent are female. Regarding ethnicity, White, Black, Hispanic, and Asian are 70, 7, 18, and 2 percent, respectively. 48 percent of my sample is eligible for free and reduced-fee lunch, while almost 12 percent are eligible for special education. While nearly 13 percent of the students are enrolled in ELL classes in kindergarten, less than 1 percent are homeless in kindergarten. My sample is evenly distributed around

[2]The dataset is not balanced, and I cannot track all students in all the grades.

the cutoff point, as indicated in Table 1.

I assign each child the kindergarten cutoff applicable in the year they reach the age of five. I compute 'waiting to enter kindergarten', the key explanatory variable based on the information if the student's entry is delayed from the year they turn five to the year they turn six. The primary factors I am focusing on are the children's achievement score in math and reading administered at each grade from three to eight. My additional potential outcome variables encompass indicators such as whether a child was placed in a gifted program, held back in a grade, or diagnosed with different disabilities. I standardized scores based on the assessment system, grade, and year. It is important to note that during the period of my analysis in Nebraska, the assessment system shifted from NeSA (Nebraska State Accountability test) to NSCAS (Nebraska Student-Centered Assessment System). As per NeSA, the scale scores fall within the range of 0 to 200. Scores below 84 are considered below-standard performance, scores between 85 and 135 indicate standard performance, and scores above 135 indicate above-standard performance. The change occurred starting from the 2017-2018 academic year. NSCAS uses different scoring scales for math and reading. Math scores fall between 2220 and 2890, while reading scores range from 1000 to 1550. NSCAS offers alternate NSCAS for special education. According to NSCAS, English proficiency is determined based on the ELA (English Language Arts) score, which consists of three components; reading comprehension, reading vocabulary, and writing. Due to the unavailability of separate sub-categories, I will consider their overall ELA score to gauge their reading proficiency. I measure indicators of being gifted by assigning one if a student is selected in the gifted program within grade four; otherwise, zero. I do not consider a student as gifted if they enter the program after grade four, as there are very few cohorts available at upper grades. Within my sample, approximately 15 percent of students have been identified as gifted. However, following the implementation of the policy reform, there was a slight decrease in the rate of gifted identification. My measure of grade retention is an indicator that a student has repeated any grades from kindergarten to grade four. In this sample, very few students are repeating the grade. I define an indicator for special education, which is set to one if a student is ever diagnosed with any form of disability by grade four. Almost 23 percent of the sample are getting special education. The summery statistics for outcome variables are presented in Table A1.

Table A2 presents summary statistics of mean achievement scores at specific points of the score distributions. It depicts the mean and standard deviation for both standardized math and standardized reading scores along with the individual level controls. Notably, individuals within the lowest 25 percent of the score distribution exhibit a higher likelihood of meeting criteria for the Free or Reduced Lunch program (FRL-eligible), identifying as Black or Hispanic, and residing in metropolitan areas (which, in this context, include Lincoln and Omaha). Conversely, for those within the top 25 percent of both score distributions, there is a probability exceeding 80 percent that they are White. This suggests a variance in the distribution across different percentiles, showcasing some degree of heterogeneity.

3 Kindergarten Cutoff, Waiting, and Test Scores

In this section, I will delve into the rationale behind my research design. I aim to explore how a child's birthday position relative to the cutoff date impacts the likelihood of waiting and, in turn, influences test scores.

3.1 The First Stage

Figure 1 shows the relationship between birthdates throughout the year and probability of treatment. I define treatment following the definition of Ricks (2022) and Barua & Lang (2016). Given, the family have the option to enroll the child in kindergarten when they reach the age of five, or they can choose to wait until the following year. I use the decision of 'wait to enter kindergarten' as the treatment.

The treatment of 'waiting to enter kindergarten' offers several advantages compared to other approaches. Numerous studies have employed various modes of age as treatments, such as predicted age (Elder & Lubotsky, 2009), relative age (Peña, 2017; Thompson et al., 2004), entry age (Angrist & Krueger, 1992; Shapiro, 2023; Fertig & Kluve, 2005), among others. In contrast, my treatment avoids the ambiguities associated with using entry age or predicted age as a treatment. Furthermore, disentangling different age effects, like relative age and absolute age, is not easily achievable. By using 'wait to enter kindergarten' as treatment, there is no need to separate these age effects (Ricks, 2022). Moreover, this treatment does not violate the monotonicity assumptions, which other age based treatments do not satisfy (Barua & Lang, 2016).

The kindergarten cutoff in Nebraska had always been October 15 before the enactment of new law. Children born after October 15 were required to wait an additional year before starting kindergarten. The probability of waiting jumps from 0.53 to 0.98 right after October 15. This pattern is also evident in Figure 1(b) after the policy change. Interestingly, since the 2012-13 academic year, Nebraska has two cutoffs for kindergarten enrollment. The cutoff was moved from October 15 to July 31, and at the same time, the restrictions on early entry were eased by allowing assessment-based enrollment for children born between these two cutoff dates. Consequently, the probability of waiting now jumps at July 31, although the increase is less than 1, indicating the prevalence of noncompliance on both sides of the cutoffs. However, the likelihood of waiting still moderately increases at October 15. Based on the magnitude of the jump in the probability of waiting at both cutoffs, I can conclude that in Nebraska, most parents tend to adhere to the kindergarten entry policy and the recommendation by waiting until the following Fall if their children's birthday falls after the soft cutoff, which is now July 31.

3.2 The Reduced-Form Relationship

Figure 2 depicts the reduced-form relationship between children's date of birth and their standardized test scores. This connection mirrors the pattern seen in the probability of waiting, as shown in Figure 1. In Panel A, we observe this relationship prior to the policy change. When the cutoff was October 15, the oldest students in the class were born in late October or November. Similarly, the youngest students were born in the first part of October. Math test scores remain relatively consistent until mid-October, after which there's a noticeable increase in math scores. Students born just after the October cutoff exhibit a math score that is 0.11 σ higher than their peers born just before the cutoff. Similar patterns can be observed for reading score. Conversely, under the new kindergarten entry policy with two cutoffs, one on July 31 and another on October 15, a distinct trend emerges. A significant surge in math scores is observed at the July cutoff point. However, at the second cutoff in October, there is a decline in math scores. The largest benefit from waiting is observed for those born after July 31 and who are recommended to wait. The advantage of waiting is somewhat reduced for those born after October 15, although the impact of waiting is still notable compared to those who do not need to wait. The magnitude of impact is almost identical for reading scores. These reduced-form relations do not account for noncompliance, which leads to a bias in the magnitude of discontinuity around cutoff points, tending towards zero (Attar & Cohen-Zada, 2018).

4 Empirical Strategy

Because of the concerns arising from manipulation of entry age, historically kindergarten entrance age literature uses IV estimates to capture the impact of entrance age. Predicted age (Elder & Lubotsky, 2009), number of days between children' 5th birthday and kindergarten cutoff date (Datar, 2006), and quarter of birth (Angrist & Krueger, 1992) are some of the instruments in the literature on kindergarten cutoff. However, any instrument that uses variation in month or quarter of birth will not satisfy the monotonicity assumption for the identification of LATE (Barua & Lang, 2016). Because of this, I use a binary instrument equals to one if the student is required by the state law to wait an additional year to start kindergarten or zero otherwise following the work of Barua & Lang (2016).

My goal is to estimate the return from waiting an additional year to enter kindergarten versus not waiting. OLS cannot yield unbiased estimate as parents can select into the treatment. The randomly assigned cutoff may solve the problem of selection bias. As the treatment effect is only identifiable at cutoff, this RDD approach produces Local Average Treatment Effect only.

4.1 Impact of Waiting on Achievement Scores: RD Approach

The kindergarten entry policy provides a natural framework for Regression Discontinuity Design. Due to con-compliance on both sides of the cutoff, RDD takes fuzzy form. I estimate waiting effects by comparing achievement scores of students just below the kindergarten cutoff to scores for students just above the cutoff. The intuition behind using this strategy is that students with birthdays very close to the cutoff, on both sides, are comparable in terms of observable and unobservable characteristics. The estimates I obtained are known as the local average treatment effect for children whose admission decision depends on which side of the cutoff his birthday stands, namely the group of compliers. The assigned variable for treatment, D_i takes the value of one if he starts kindergarten at age 6 or more and zero otherwise. I estimate the following equation for educational outcomes at cut-off:

$$Y_i = \alpha + \beta D_i + \tau X_i + \epsilon_i \tag{1}$$

where Y_i is the educational outcomes for individual i measured, D_i is an indicator which takes the value one if the individual i's kindergarten entry is delayed from the year when he turns five to the year when he turns six, X indicates personal characteristics, including sex, race. β captures the effect of waiting an additional year to start kindergarten on educational outcomes holding ϵ_i and X_i constant. The estimator in the model mentioned above might be biased because parents might influence D by speeding up or red-shirting their children's school entry. To overcome this, I project Y_i and D_i as follows:

$$Y_i = \mu Z_i + \pi(birthday - cutoff) + \delta Z_i(birthday - cutoff) + u_i \tag{2}$$

$$D_i = \theta Z_i + \sigma(birthday - cutoff) + \eta Z_i(birthday - cutoff) + v_i \tag{3}$$

where Z_i characterizes incentives behind the delay in kindergarten entry for individual i. I defined Z as below:

$$Z = \mathbb{1}[r > Cutoff] \tag{4}$$

where r indicates running variable, which is in this case 'date of birth'. I consider cutoff to be 15th October for the years before policy change and 31st July after reform. Due to the variability in cutoffs, I normalize cutoff around zero and adjust the running variable accordingly. Z captures whether a student is recommended to delay kindergarten entry. Here, $\mu = \theta\beta$ by linear projection. Under smoothness of the running variable and assuming $\theta \neq 0$, β is identified as the ratio of discontinuity at Y_i to that of D_i.

While estimating equation (2) and (3), I restricted my sample to children with birthday within a relatively narrow window around the kindergarten cutoff to avoid influence of children whose birthday is far from the cutoff. I use 45 days of data bandwidth for analysis. The traits of individuals within 45 days around the cutoff are highly similar. The estimates are not sensitive to the choice of bandwidth. For this design to be effective, it is crucial that Z is not correlated with unobserved characteristics. This implies that Z influences Y solely through D within the specific narrow window around the cutoff. This estimate captures average treatment effect among children whose school entry decision gets induced by the instrument.

4.1.1 Identification

To produce consistent estimates, the identifying strategy requires some conditions to hold. First, the monotonicity condition is needed to be satisfied. The monotonicity assumption requires that crossing the kindergarten cutoff does not make an individual less likely to wait an additional year to start kindergarten. What implies is that if a child waits an year when he is born before the cutoff, he would also wait when his date of birth is just after the cutoff. My setting seems to ensure no presence of defiers in the sample.

Second, date of birth around the cutoff cannot be manipulable. If the parents manipulate date of birth to expedite the school entry of their children, distribution of running variable at the cutoff will be discontinuous. Since parents are unlikely to determine the exact date of birth, children born before and after the cutoff are similar in terms of observable and predetermined characteristics (McCrary & Royer, 2011). However, there is possibility of differential selection into public school before and after the cutoff. If there is tendency among parents to use private schools as a means to bypass the entrance age requirement (Taveras, 2021) of school entry policy, the condition of continuity at the cutoff might be violated. To address this concern, I check the evidence of discontinuity at the cutoff. I run rddensity test that uses a local cubic estimator to find the evidence of discontinuity (Cattaneo et al., 2018). This test fails to provide any statistically significant evidence of discontinuity around the cutoff (Figure A1). The density test's result suggest a very similar admission rate in public schools of Nebraska among children born before and after the cutoff. The distribution of date of births is also consistent with the above findings. Figure 3 displays a histogram depicting the birthdates of all the children, and it appears to exhibit a smooth distribution near the cutoff point.

Third, to ensure the validity of RD estimates, it is recommended to test the balance of observable covariates around the cutoff. The goal is to assess whether the observations just above the cutoff are similar to those under the cutoff. In Table 2, I provide estimates obtained from the test of discontinuities in predetermined covariates by applying both parametric and non-parametric approaches. In both cases, I estimate if there is any discontinuity in characteristics around the cutoff. I employ bandwidths of 45 and 30 days for each covariate. Using 45 days of bandwidth, I could not find any evidence of discontinuity for any covariates. However, with a 30-day bandwidth, I observed only one significant jump at the threshold for the black covariate. The findings are also reflected in Figure 4. In my sample, I do not have children whose parents utilize private schools to bypass the kindergarten threshold and then transition their children back to public schools in subsequent grades.

There is another visual test for manipulation which can be used when there might be some discontinuity in the runny variable for exogenous reasons. According to Lee & Lemieux (2010), a sufficient condition for an unbiased RD estimates requires smoothness of conditional joint distribution of the running variable.

Unless other approaches which require only the continuity of running variable, this test requires the ratio of conditional and unconditional densities of the running variable to be continuous. If there is some discontinuity in the running variable due to some factors exogenous to the treatment determination, that discontinuity needs to be balanced to make the ratio continuous (Zimmerman, 2014).

Figure A2 shows density ratios for six groups: White, Black, Hispanic, female, FRL eligible students, and students living in metropolitan areas. Each point indicates the ratio of the proportion of children with specific characteristics to the proportion of all observation within the specific bin. In accordance with a valid RD design, each density ratio remains continuous around the threshold.

4.2 Impact of Waiting on Achievement Scores: Quantile Approach

To estimate the effect of waiting on the distribution of achievement score at each points in the score distribution other than mean, I need to apply quantile regression framework. At the θth quantile, the basic quantile model can be expressed as:

$$y_{i\theta} = \beta_\theta D_i + \epsilon_{i\theta} \tag{5}$$

where $Y_{i\theta}$ denotes the θ quantile of Y given D. β_θ is the parameter of interest which indicates the coefficent of interest at θ quantile. It captures the difference in the conditional θ quantiles of outcomes Y_1 and Y_0. This model is estimated by minimizing the following equation according to Koenker & Bassett Jr (1978) as follows:

$$min \sum_{i:y_i>=\beta_\theta D_i}^{N} \theta|y_i - \beta_\theta D_i| + \sum_{i:y_i<\beta_\theta D_i}^{N} \theta|y_i - \beta_\theta D_i| \tag{6}$$

The value of the coefficient, β will vary based on the specific quantile being studied. The QTE represents the impact of treatment D on different points of the marginal distribution of potential outcomes. As D is endogenous in this case, the conditional quantile of outcome given the treatment is not equal to the quantile of potential outcome (Chernozhukov & Hansen, 2004). Conventional quantile regression fails to produce an unbiased estimate and calls the need for quantile regression with instrumental variables. To solve this endogeneity problem, I have utilized the instrumental variable quantile regression (IVQR) model proposed by Chernozhukov & Hansen (2004). The model requires instrument to affect outcomes only through the treatment. Under some conditions, IVQR model satisfies the following conditional probability:

$$Pr\{y - D\beta_\theta <= Z' * 0|Z\} = \theta \tag{7}$$

β_θ should be estimated in such a way that the coefficient of Z gets very close to zero. In particular, I use smoothed estimating equations (SEE) outlined in Kaplan & Sun (2017). This estimator applies kernel method to smooth the original estimating function. Moreover, this estimator facilitates the testing of certain hypotheses related to the endogenous variable.

Using IVQR framework is justified when impacts demonstrate evidence of heterogeneity across the distribution. It can be tested with a post-estimation command used after SEE estimation. That test supports the adaptation of the quantile approach[3](Table A4).

5 Results

5.1 RD Approach

This section presents the effect of waiting an additional year on educational outcomes. The regressions are estimated on the sample of students of public schools in Nebraska. I start with providing evidence that, children born within 45 days before the cutoff are similar to the ones born after 45 days of the cutoff. Table A3 presents prevalence statistics for six covariates; namely, female, FRL eligible,

[3]If Kolmogorov–Smirnov statistic is greater than the critical value, we can reject the null hypothesis. The null hypothesis is contact effect throughout the distribution.

metro, White, Black, and Hispanic, demonstrating the evidence that 'before' and 'after' groups are close to each other and it can be expected that children are assigned at random to these two groups. Table 2 also provides additional evidence to support of using 45 days of bandwidth for analysis.

Table 3 illustrates the impacts of waiting an additional year to start kindergarten on standardized math scores, spanning from grade three to grade eight, considering the bandwidth equal to 45. In the first column, I present the results obtained through Ordinary Least Squares (OLS) analysis, which indicate a substantial and favorable impact of waiting a year to start kindergarten on test scores. Specifically, the coefficient of 0.457 suggests that waiting a year leads to an improvement of 0.457 σ in math scores. As the decision to wait suffers from endogeneity issues, I address that concern using 2SLS strategy. In the second column, I present the results from the 2SLS estimation without any additional controls. However, as demonstrated in Table A2, it becomes apparent that high-performing students often come from affluent backgrounds and are predominantly of white ethnicity. Therefore, I consider the inclusion of controls for these demographic characteristics to potentially regulate the impact of waiting a year to start kindergarten. In the third column, I provide the 2SLS estimates with individual-level controls, and I go further by accounting for factors that may remain consistent within a school district. This is achieved by introducing school district fixed effects in the fourth column. Column 4 presents the 2SLS estimates with both individual-level controls and school district fixed effects. Notably, the 2SLS estimates remain largely consistent across columns 2, 3, and 4. This suggests that the effects of delaying kindergarten are not strongly correlated with observable characteristics of the children and the school districts they attend. However, in contrast to the 2SLS estimates, OLS estimates demonstrate a more substantial positive impact of waiting on test scores. Comparing OLS and 2SLS estimates, respectively, from column (1) and (2) confirms the evidence of positive selection on gains, implying students who are most likely to wait would have more significant treatment effects, which is contrary to the findings of Elder & Lubotsky (2009). The relationship between waiting a year and achievement scores gradually diminishes over time as students progress through the grades. For math score, the decision to wait improves the test score by 0.341 σ in grade 3. By the time students reach the eighth grade, the positive effect reduces to 0.256 σ. This same pattern holds true for reading scores. However, the influence of delaying school entry does not fade entirely. Even in the eighth grade, this influence remains substantial and statistically significant for both math and reading scores.

5.2 Heterogeneity Analysis

To address the research question regarding potential variations based on gender, race, or socioeconomic status, I employed the same fuzzy regression discontinuity design without any additional controls. This allowed me to estimate Local Average Treatment Effects (LATE) for distinct subgroups. These subgroups include male and female participants (categorized by gender), individuals from metropolitan and non-metropolitan areas (classified by metropolitan status), those eligible and ineligible for Free or Reduced-Price Lunch (FRL) programs (based on socioeconomic status), and participants from different racial backgrounds such as white, black, and Hispanic. I used equations 2 and 3 to derive these estimates. Figure 5, 6, 7, and 8 display plots containing estimated coefficients and their corresponding robust confidence intervals. I arranged the results side by side for easier comparison. Upon examining the magnitude of the impact, I observed that waiting an additional year tends to have a greater effect on improving girls' test scores compared to boys. The diminishing effect of waiting is particularly noticeable among boys, where the fade-out occurs more rapidly. Conversely, for girls, the fade-out rate is considerably slower. Children from both metropolitan and non-metropolitan areas show similar responses in terms of the magnitude of impact and the fading out of this impact over time. However, the impact varies significantly across different socioeconomic statuses. Notably, the effect is more pronounced among children from affluent backgrounds, and the benefit of an additional waiting year diminishes quickly for students eligible for Free or Reduced-Price Lunch, aligning with the findings of Elder & Lubotsky (2009). While I compare the estimates across various races, I find that the response to entry age is relatively stronger among white children and advantage of an extra year fades out at a faster rate for black and Hispanic children. This observation is harmonious with the pattern seen in the comparison between Free or FRL eligible and FRL ineligible groups. The differential rate in the pace of fading out imply strong presence of Matthew effect in age at entry literature. I also examine whether the variation in impact is statistically significant or not and present the results in Table 4. I observed that gender, FRL eligibility, and race to significantly contribute to the heterogeneity in impact. When evaluating the impact of waiting while interacting with the white category, I exclude 'Asian' from the analysis. Since Asians exhibit the highest average scores, considering them as non-White, would not provide an accurate representation of the true impact of waiting and could potentially bias the results downward (A5). However, the effect is consistent for both math and reading scores.

5.3 Quantile Approach

The RD estimates of section 5.1.1 confirm the significant impact of waiting on achievement scores. The impact could potentially be more pronounced for certain students while being less pronounced for others. I therefore investigate the heterogeneity of the effect of waiting on scores using quantile regression. Specifically, to address the endogeneity of treatment variable, I employ the IVQR model. In this section, I present the causal impact of waiting to enter kindergarten on the distribution of achievement scores. The results obtained from OLS, two-stage least square (2SLS), ordinary quantile regression (QR), and instrumental variable quantile regression differ from each other.

Panel A and B of Table 3 show OLS estimates for math and reading scores, respectively. Waiting for an additional year before entering kindergarten results in a math score increase of 0.457 σ, and a reading score increase of 0.452 σ. These point estimates indicate score gain compared to those who did not wait. However, it's worth noting that OLS estimates might overstate the actual causal impact of waiting on standardized scores. To address this concern and the potential influence of other factors, I've examined the Two-Stage Least Squares (2SLS) model. In this model, the estimates for the effect of waiting are approximately 30% smaller than the OLS estimates, although they remain statistically significant.The OLS and 2SLS estimates both provide conditional mean effect of waiting an additional year to start kindergarten on achievement scores, although the distributional aspects of those estimates remain undiscovered. Quantile regression estimates of Table 5, on the other hand, provide OLS-like estimates at various points along the distribution. In this analysis, waiting for an additional year before beginning kindergarten is associated with a positive and statistically significant impact on scores across all quantiles. The magnitude of the impact varies, ranging from an increase of 0.347 σ to 0.527 σ for math scores and from 0.379 σ to 0.481 σ for reading scores. As the median point estimate surpasses the mean estimate, the distribution of both math and reading scores skews towards the left.

In Panel B of Table 5, I instrument for waiting an additional year to start kindergarten using the indicator of whether the birthday falls after the kindergarten cutoff. Whereas the 2SLS estimator captures the mean effect of waiting, IVQR provide 2SLS type estimates at various points of the score distribution. Consistent with the mean effect between OLS and 2SLS, the estimates from IVQR model are considerably smaller than their OLS counterparts. The IVQR estimate at 90th percentile of math score distribution is 0.358 σ, whereas the corresponding QR estimate is 0.537 σ. Figure 9 demonstrates the causal effect of waiting on scores by plotting estimates from QR and IVQR model. QR estimates in (i) of both Panel A and B of Figure 9 show significant variability throughout the score distribution.

The OLS estimates are only consistent with QR estimates around 30th and 20th percentile for math and reading, respectively. Below that range point estimates are much smaller than the OLS estimate in magnitude. On the other hand, after that range estimates are significantly larger than the OLS estimate. (ii) of Panel A and B present the IVQR estimates. Below 25th percentile the IVQR estimates are smaller than the normal 2SLS estimate. IVQR estimates are harmonious with the 2SLS estimate from 25th to 70th percentiles. Beyond that point, IVQR outgrows the 2SLS estimate. However, at higher quantiles, due to fewer observations, standard errors widen and make the IVQR estimate not significantly different from the 2SLS estimate. The IVQR estimates outgrow 2SLS between approximately the 70th and 85th percentiles. The pattern of IVQR estimates is quite similar for reading scores as well.

6 The Effect of Changing Kindergarten Cutoff on Achievement Gap

After the enactment of No Child Left Behind in 2001, more emphasis has been put on the improvement of test scores and accountability like never before. To ensure better performance in inter-state comparison of scores, many states are moving the kindergarten cutoff earlier (Cannon & Lipscomb,2008). In the fall of 1968, 96 percent of six-year-old and five-year-old were enrolled in 1st grade and kindergarten, respectively, or above. By 2005, the ratio has fallen drastically to 84 percent. Following this trend, Nebraska has increased the minimum kindergarten entrance age by 2.5 months. Moving the cutoff earlier in the calendar year increases the school entry age by making more students wait. Though entrance age effect is positive among students from lower socioeconomic status families, the overall waiting a year might exacerbate socioeconomic achievement gap in academic performances.

In Section 5.2, I have identified variations in the effects of waiting on various groups of children, such as male-female, FRL (Free and Reduced Lunch) eligible-FRL ineligible,and White-non White. I find the impact of waiting an additional year to start kindergarten to be larger among children from non-poor families. Moreover, quantile IV regressions assure the stronger impact among students from the upper quantile of score distribution. As there are more prevalence of students from high income families on the top quantile, it confirms heterogeneity in impact across socioeconomic groups. This disparity in impact raises another research question: whether the policy reform amplifies or diminishes the achievement gap among these different groups.

To investigate how advancing the kindergarten cutoff date impacts the socioeconomic achievement gaps, I have estimated the following equation:

$$Y_{it} = \alpha + \beta FRL_i + \gamma post_t + \delta post_t x FRL_i + \eta_{it} \tag{8}$$

In this equation, Y_{it} represents the test score of an individual i at time period 't'. FRL_i is an indicator that signals whether the individual is eligible for Free and Reduced Lunch (FRL). Additionally, $post_t$ is an indicator that determines whether the year of kindergarten admission falls before or after the change in policy. The parameter of interest, represented by δ, indicates whether the policy reform improves or deteriorates the achievement gap.

This research design identifies the effect of changing kindergarten cutoff on achievement gap under the assumption that there exists a parallel trend in test scores between FRL eligible and FRL ineligible groups of students before the policy reform using difference-in-difference method. The first difference comes from the gap between FRL eligible and FRL ineligible group. The second difference comes from comparing test scores before and after the policy change.

My results from Figure 10 demonstrate that following the policy change, the average standardized math scores improved for all children, irrespective of their FRL eligibility status. However, the rate of improvement in scores doesn't differ significantly between these groups. Altering the kindergarten cutoff appears to neither enhance nor exacerbate the existing achievement gap. To delve deeper into the dynamics, I break down the entire year span into intervals based on how children's date of birth is influenced by the policy change. To facilitate comparison, I conduct the following regression:

$$Y_{it} = \alpha + \beta FRL_i + \gamma post_t + \pi direct_i + \delta post_t x FRL_i + \mu post_t x direct_i$$

$$+ \rho FRL_i x direct_i + \lambda FRL_i x direct_i x post_i + \eta_{it}$$

(9)

Here, *direct* signifies whether a child's birthday falls between Aug 01 and Oct 15. Children born within these dates are the ones whose eligibility for kindergarten admission is directly impacted by the policy reform. If a child is born outside these dates, the eligibility condition is not directly influenced, and the *direct* variable is assigned a value of zero for them. The coefficient of interest, denoted as λ, indicates whether the policy change is advantageous for the directly affected children who are FRL eligible.

Figure 11 illustrates that shifting the kindergarten cutoff to an earlier date has a similar impact on the socioeconomic achievement gap as shown in Figure 10. However, it does reduce the achievement gap in reading scores specifically for children born between the cutoffs, whose kindergarten entry is directly influenced by this policy change.

7 Robustness Checks

Figure A3 illustrates that the results are not sensitive to the selection of bandwidth. Across the range from 15 to 75, the estimated coefficient remains largely unaffected by the choice of bandwidth. I also perform an additional falsification test by examining for any discontinuity at various dates around the cutoff. However, I do not find any evidence of a discontinuity in the outcome variable at those randomly selected dates (Figure A5).

There is an concerns regarding potential data attrition. To investigate this, I predicted the probability of missing exam scores. The results in Table A6 show that crossing the threshold has no association with missing values of exam scores until grade 6. However, starting from grade 6, there is a noticeable reduction in the likelihood of missing test scores at the cutoff. It's essential to highlight that these estimates, although statistically significant, are of a minimal magnitude. Despite these findings, the results appear to be relatively unresponsive to data attrition. In Figure A4, I present the results while considering this attrition issue. The base category and another category, which only includes cohorts with scores from all grades, produce almost identical estimates. However, the estimates from the balanced panel are slightly lower but still closely aligned with those from the base category.

8 Conclusion & Discussion

My results suggest that waiting an additional year to start kindergarten boost the achievement test scores. In particular, this delay enhances both math and reading scores by more than 0.3 standard deviations. These results align well with previous studies (Bedard & Dhuey, 2006; McEwan & Shapiro, 2008; Puhani & Weber, 2008; Peña, 2017). Although the advantageous impact diminishes over time, it remains notably stronger compared to the findings of Elder & Lubotsky (2009) in later grades and is comparable to Cascio & Schanzenbach (2016). For instance, the effect on math scores decreases from 0.341 standard deviations to

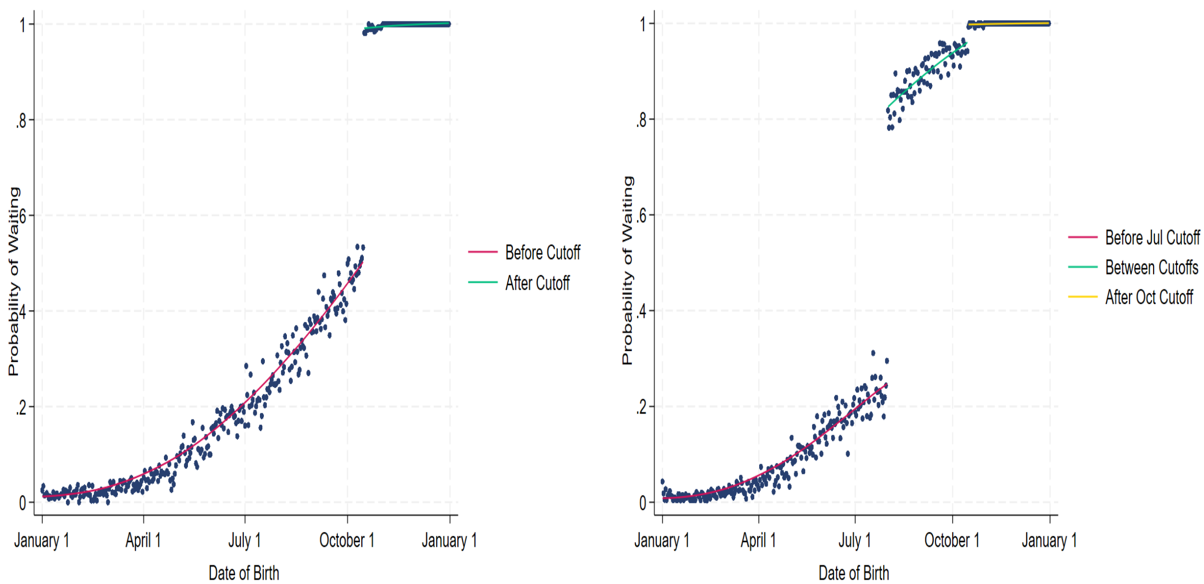
0.256 standard deviations, which is quite strong and significant. Furthermore, the observed improvement is more significant for students positioned near the top of the score distribution as opposed to those at the lower end.

It is crucial to highlight that the effects of entry age vary among individuals. I observe a more pronounced diminishing impact among students from disadvantaged backgrounds, whereas students from advantaged backgrounds experience the benefits of entry age for a more extended period. This observation aligns with the findings of Elder & Lubotsky (2009), supporting the notion of skill accumulation before starting kindergarten. Structural barriers disproportionately hinder students from disadvantaged groups from gaining the advantages of delaying entry (Ricks, 2022). If this policy change focuses only on kindergarten entry age, it contributes to a widening achievement gap as at-risk children are already lagging behind in terms of human capital. To narrow this gap, it is imperative to eliminate barriers and enhance investments in high-quality pre-kindergarten and kindergarten programs (Ricks, 2022; Cannon & Lipscomb, 2008). Considering the substantial financial burden on disadvantaged households associated with holding back children for a year, the policy has not proven successful in diminishing the achievement gap in their favor (Gelbach, 2002; Cascio, 2006).

While advancing the kindergarten cutoff date is likely to enhance academic achievement, policymakers should carefully weigh the associated costs, particularly for disadvantaged households. Instead for just delaying the entry, it is crucial to explore potential connections between school entry age, childcare costs, and the achievement gap in order to formulate a more effective and equitable policy for everyone.

9 Figure

Figure 1: 1st Stage: Before and After Policy Change

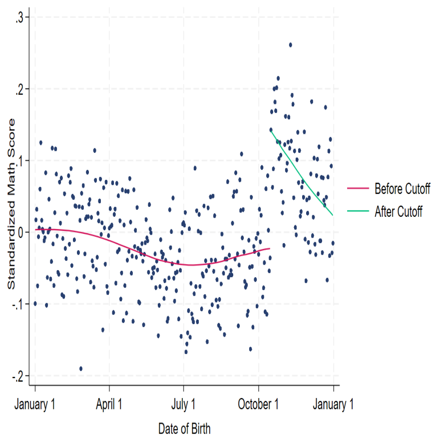


(a) Before Policy Change

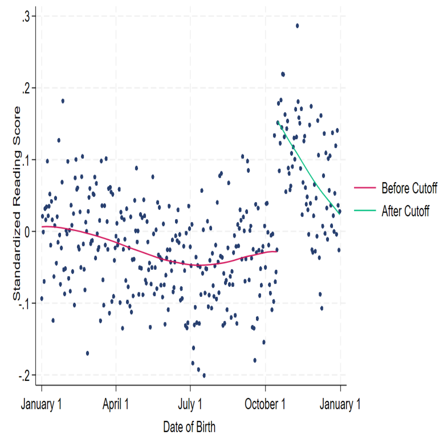
(b) After Policy Change

Notes: This figure presents pattern in waiting over date of birth . (a) displays the plot of the first stage prior to the policy reform, whereas (b) represents the plot after the reform. Both scatter plots illustrate the likelihood of delaying entry into kindergarten based on birthdate, along with their corresponding best-fit lines. In both the graphs points are daily averages. Waiting is defined as one if an individual's school entry is delayed from the year in which he turns 5 to the year he turns 6. The sample consists of all the students who have started kindergarten within the age of 57 to 79 months, and for whom I can see the year of kindergarten entry.

Figure 2: Reduced-Form Relation: Before and After Policy Change

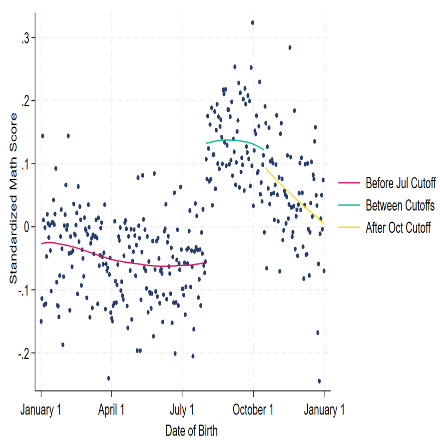


(i) Math Score

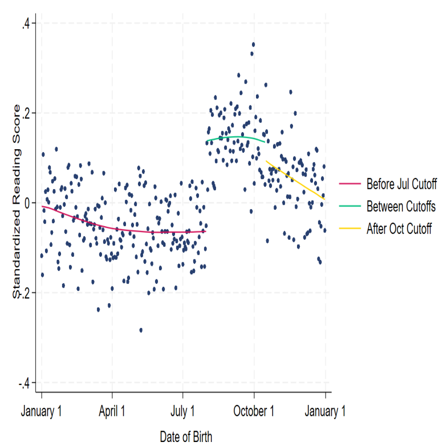


(ii) Reading Score

Panel A: Before Policy Change



(i) Math Score

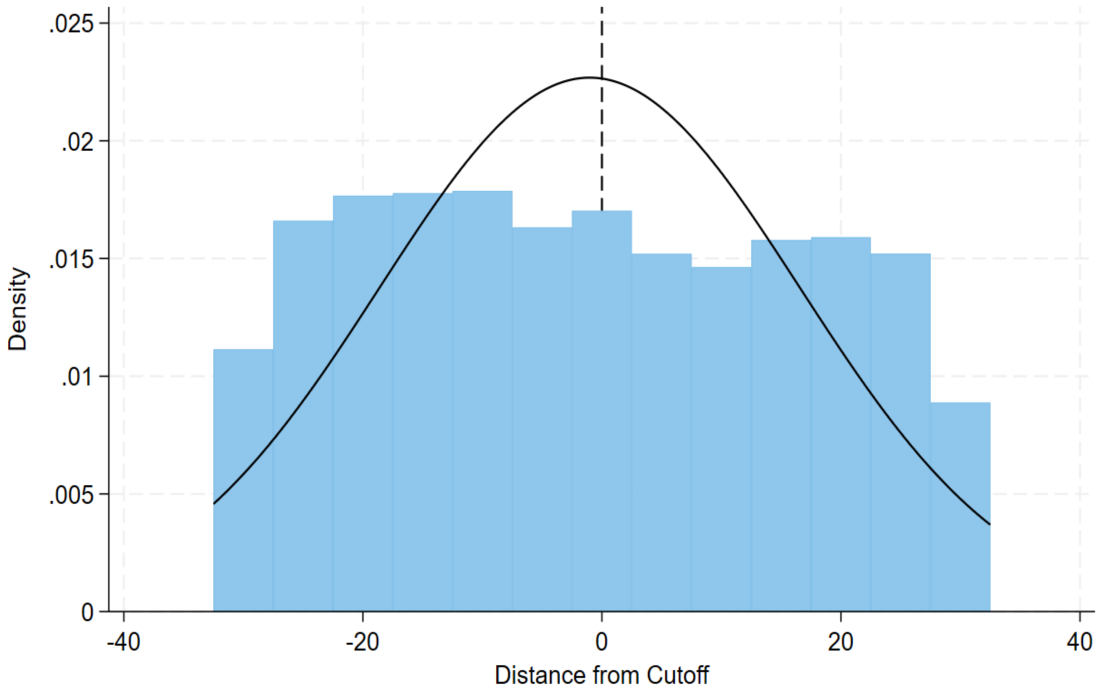


(ii) Reading Score

Panel B: After Policy Change

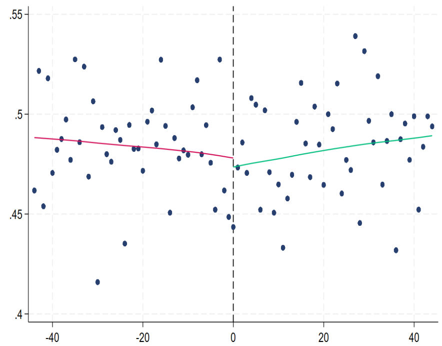
Notes: This figure presents pattern in standardized math achievement scores of grade 3 over date of birth, along with their corresponding best-fit lines. Panel A displays the plot of the reduced form prior to the policy reform, whereas Panel B represents the plot after the reform. In both the graphs points are daily averages. All the scores are standardized by assessment name, grade and year. The sample consists of all the students who have started kindergarten within the age of 57 to 79 months, and for whom I can see the year of kindergarten entry.

Figure 3: Histogram of date of birth

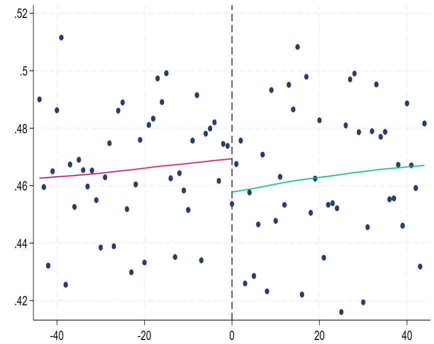


Notes: Histogram of date of birth of sample students born 30 days around the cutoff. Separate columns are shown for each bin where bin width is 5 days. The sample consists of all the students who have started kindergarten within the age of 57 to 79 months, and for whom I can see the year of kindergarten entry.

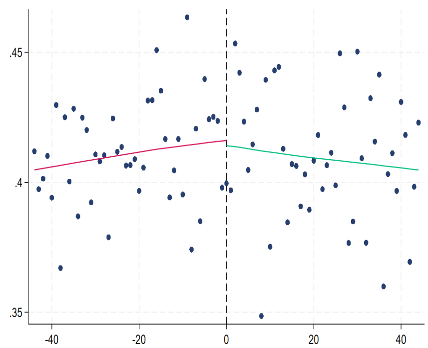
Figure 4: Covariate Balance Test



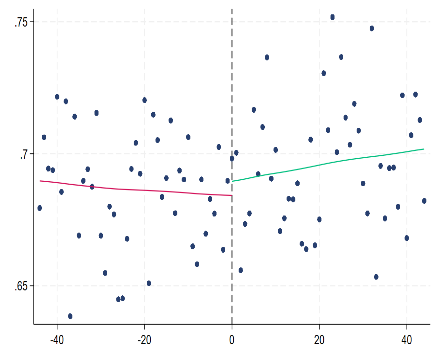
(a) Female



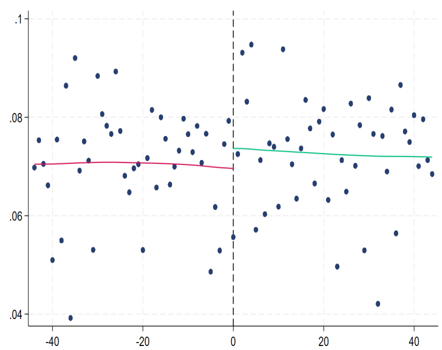
(b) FRL Eligible



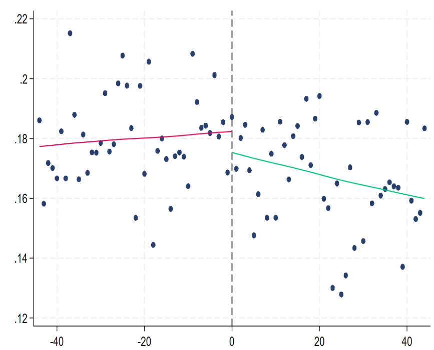
(c) Metro



(d) White



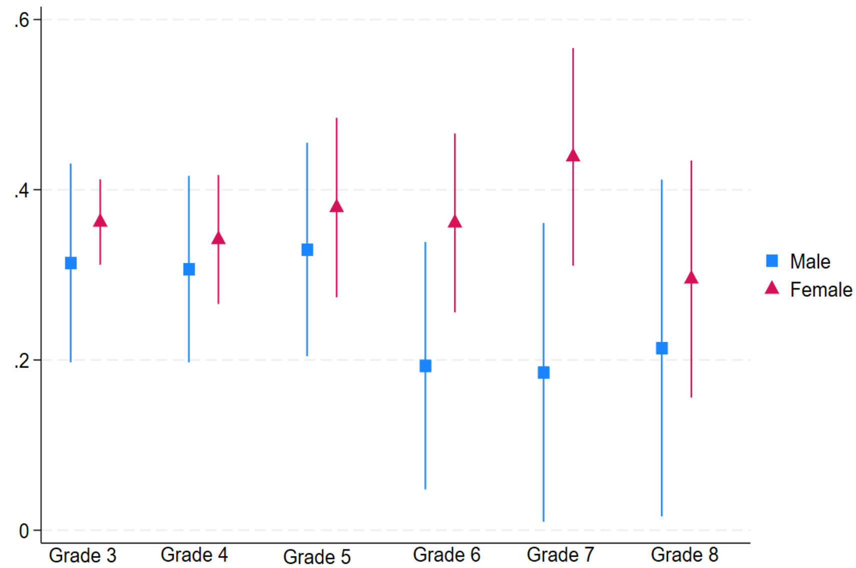
(e) Black



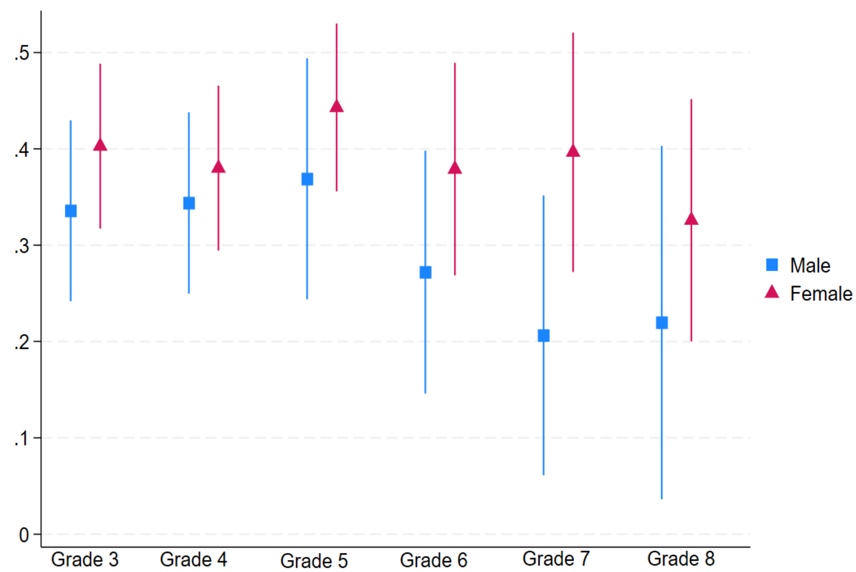
(f) Hispanic

Notes: Scatter plots represent mean of demographic characteristics by distance relative to the cutoff. 45 days of bandwidth has been taken. All points indicate daily average of relevant covariates. Lines are local polynomial smooth plot on fitted values.

Figure 5: Estimates by Grades: Gender wise



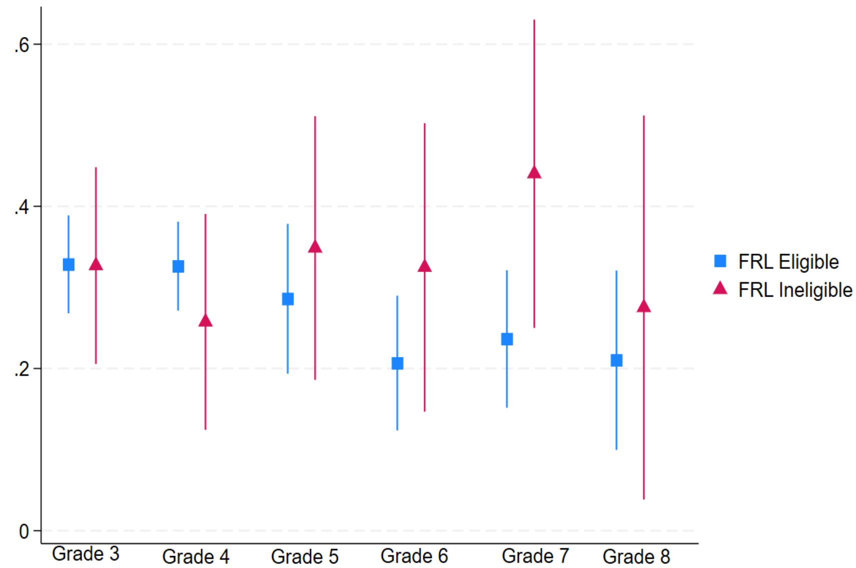
Standardized Math Score by Grade



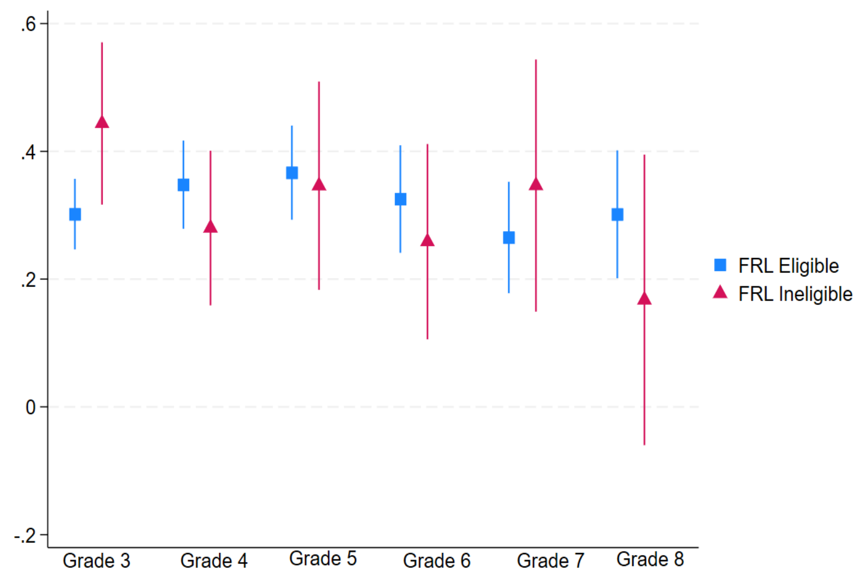
Standardized Reading Score by Grade

Notes: All treatment effects were estimated separately for boys and girls using 2SLS specification considering 45 days of bandwidth around the cutoff. Treatment is waiting an additional year to start kindergarten. Standard errors are clustered at school district level. All regressions are run without controls. Test scores are standardized by assessment system, year, and grade.

Figure 6: Estimates by Grades: FRL Eligibility wise



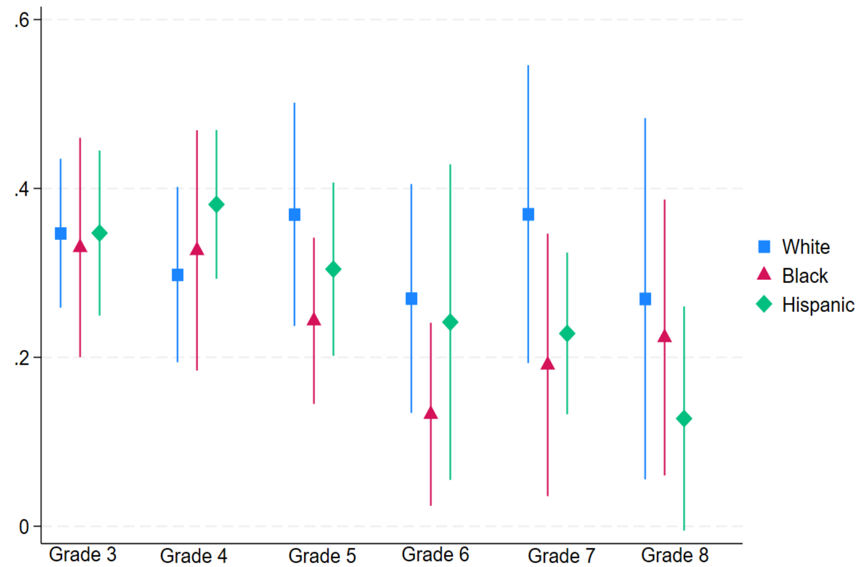
Standardized Math Score by Grade: FRL Eligibility-wise



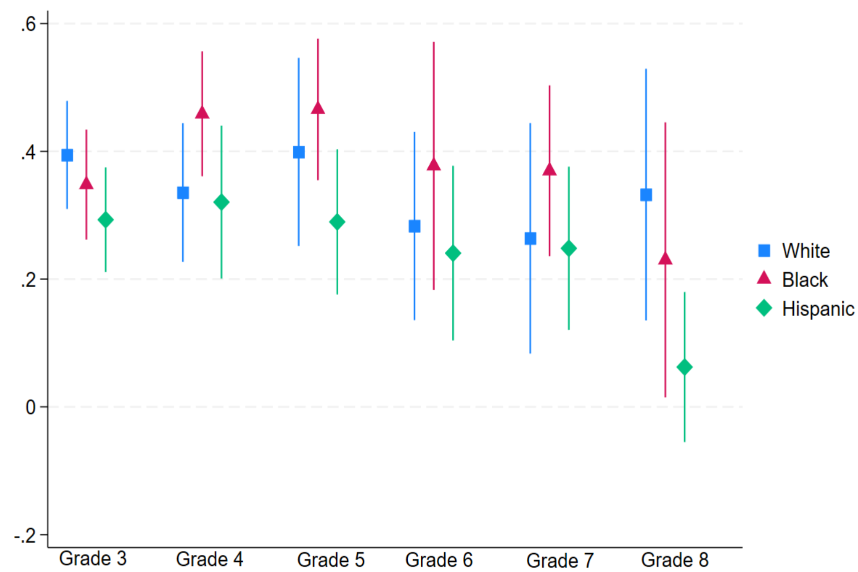
Standardized Reading Score by Grade: FRL Eligibility-wise

Notes: All treatment effects were estimated separately for FRL eligible and FRL ineligible students using 2SLS specification considering 45 days of bandwidth around the cutoff. Treatment is waiting an additional year to start kindergarten. Standard errors are clustered at school district level. All regressions are run without controls. Test scores are standardized by assessment system, year, and grade.

Figure 7: Estimates by Grades: Across Races



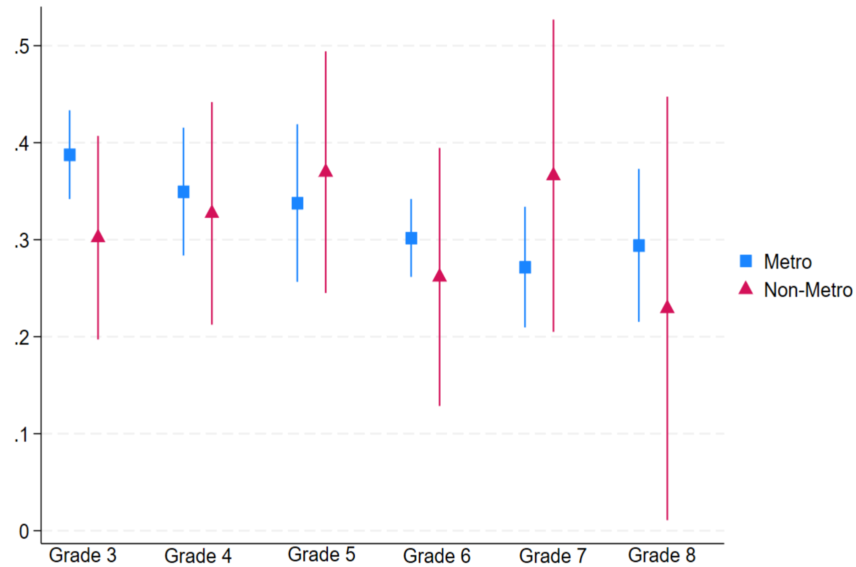
Standardized Math Score by Grade: Across Races



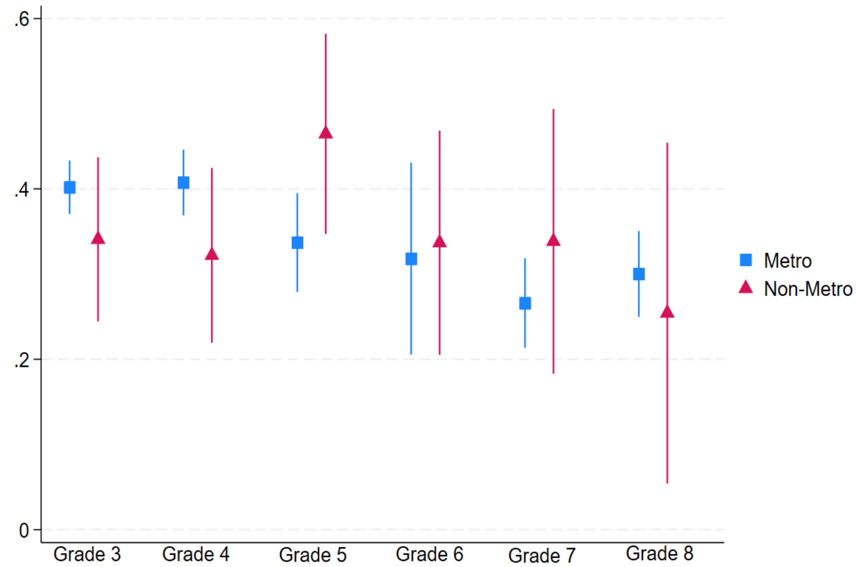
Standardized Reading Score by Grade: Across Races

Notes: All treatment effects were estimated separately for White, Black, and Hispanic students using 2SLS specification considering 45 days of bandwidth around the cutoff. Treatment is waiting an additional year to start kindergarten. Standard errors are clustered at school district level. All regressions are run without controls. Test scores are standardized by assessment system, year, and grade.

Figure 8: Estimates by Grades



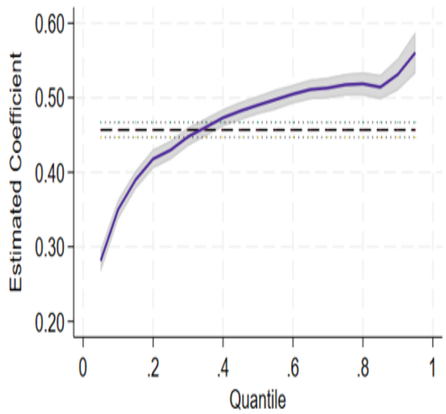
Standardized Math Score by Grade

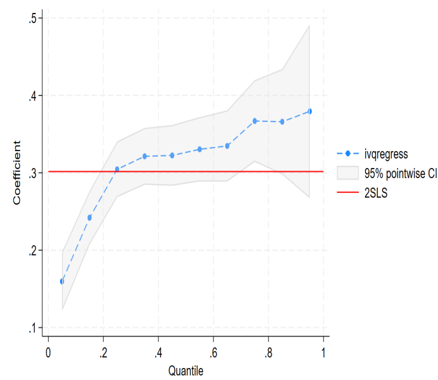


Standardized Reading Score by Grade

Notes: All treatment effects were estimated separately for students living in metropolitan area and non-metropolitan area using 2SLS specification considering 45 days of bandwidth around the cutoff. Treatment is waiting an additional year to start kindergarten. Standard errors are clustered at school district level. All regressions are run without controls. Test scores are standardized by assessment system, year, and grade. Only Lincoln and Omaha are considered as metropolitan area in this analysis.

Figure 9: Quantile Regression Estimates

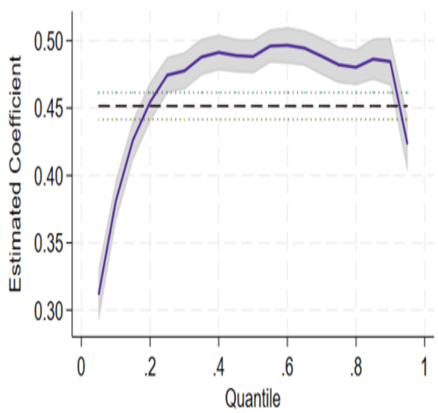


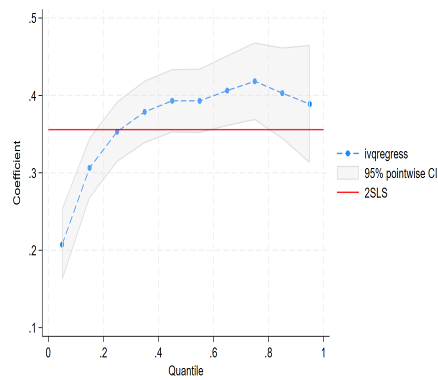


(i) QR Estimates

(ii) IVQR Estimates

Panel A: Standardized Math Score





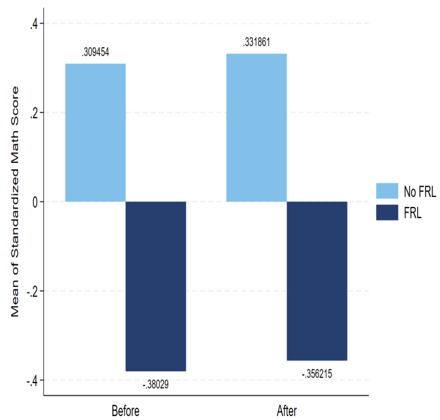
(i) QR Estimates

(ii) IVQR Estimates

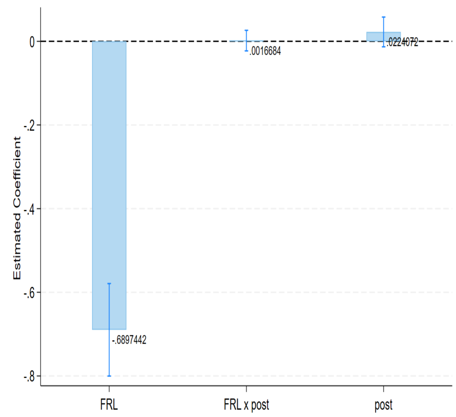
Panel B: Standardized Reading Score

Notes: QR and IVQR estimates for treatment effect on standardized test score. Coefficient estimates are on the vertical axis and the quantile index is on the horizontal axis. Treatment is waiting an additional year to start kindergarten. All regressions are run without controls. The shaded region indicates 95% confidence interval. Robust standard errors are used for inference purpose.

Figure 10: Socioeconomic gap in test scores-I

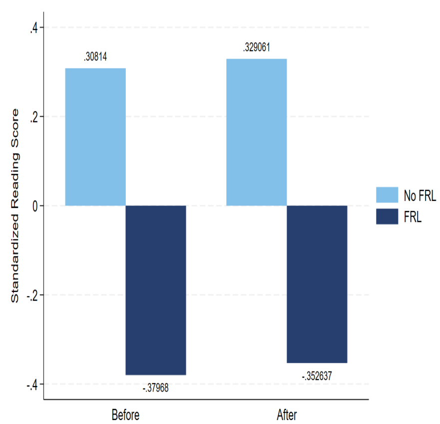


(i) Average standardized math score by FRL status

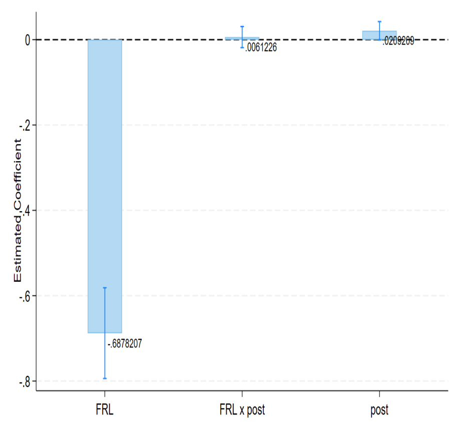


(ii) Socioeconomic differences in Math score

Panel A: Standardized Math Score



(i) Average standardized reading score by FRL status

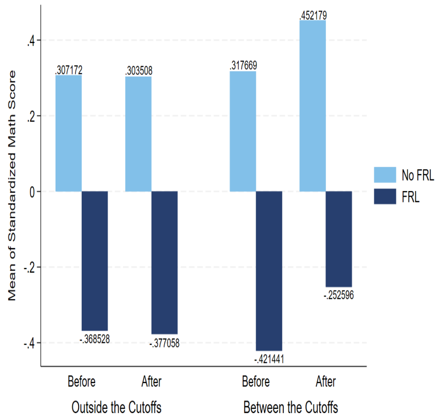


(ii) Socioeconomic differences in Reading score

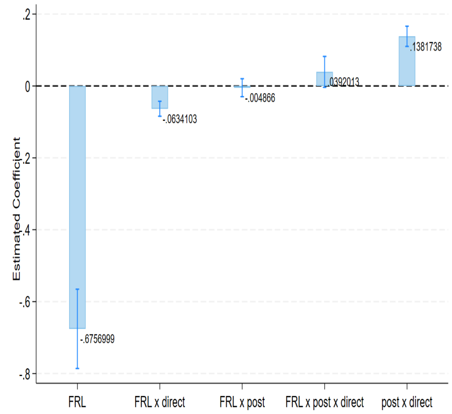
Panel B: Standardized Reading Score

Notes:(i) of both panels present the average standardized test score by FRL status. (ii) of both panels show estimates from equation (9). The shaded region of (ii) indicates 95% confidence interval. The sample size for Figure 10 is 771,143.

Figure 11: Socioeconomic gap in test scores-II

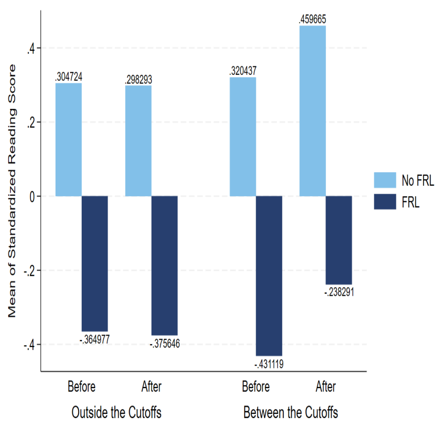


(i) Average standardized math score by FRL status and time period

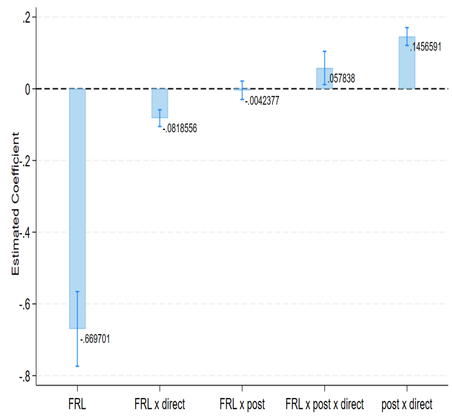


(ii) Socioeconomic differences in Math score

Standardized Math Score



(i) Average standardized reading score by FRL status and time period



(ii) Socioeconomic differences in Reading score

Standardized Reading Score

Notes:(i) of both panels present the average standardized math score by FRL status and time period. (ii) of both panels show estimates from equation (10). The shaded region of (ii) indicates 95% confidence interval. The sample size for Figure 11 is 771,143.

10 Table

Table 1: Descriptive Statistics

	Before change in cutoff		After change in cutoff	
	Before Oct cutoff	After Oct cutoff	Before Jul cutoff	After Jul cutoff
Age in months	63.74	67.82	64.05	68.39
	(4.33)	(1.09)	(3.51)	(2.66)
Female	0.49	0.48	0.48	0.48
	(0.50)	(0.50)	(0.50)	(0.50)
FRL eligible	0.46	0.47	0.48	0.47
	(0.50)	(0.50)	(0.50)	(0.50)
EL eligible	0.13	0.12	0.14	0.12
	(0.33)	(0.32)	(0.34)	(0.33)
SPED	0.12	0.11	0.13	0.12
	(0.32)	(0.31)	(0.34)	(0.33)
White	0.71	0.71	0.67	0.67
	(0.46)	(0.46)	(0.47)	(0.47)
Black	0.07	0.08	0.07	0.07
	(0.26)	(0.27)	(0.25)	(0.25)
Hispanic	0.17	0.16	0.19	0.18
	(0.37)	(0.37)	(0.39)	(0.39)
Asian	0.02	0.02	0.02	0.02
	(0.14)	(0.14)	(0.15)	(0.15)
Metro	0.41	0.41	0.41	0.41
	(0.49)	(0.49)	(0.49)	(0.49)
Observations	84,662	21,245	51,065	34,428

Notes: Sample consists of public school students of Nebraska starting kindergarten from 2007-2008 to 2015-2016. All students are included in the sample disregarding the choice of bandwidth. All the students in the sample have started kindergarten at the age of 57 to 79 months. Total number of observations is 191,400.

Table 2: Validity of the Regression Discontinuity Design

	Parametric Estimation		Non-parametric Estimation	
	BW=45	BW=30	BW=45	BW=30
Female	-0.000827	-0.00529	-0.01005	-0.00949
	(0.00930)	(0.0114)	(0.01006)	(0.01231)
FRL Eligible	-0.0101	-0.0157	-0.01637	-0.01779
	(0.00928)	(0.0114)	(0.01005)	(0.0123)
White	0.00541	-0.0103	0.00479	0.00593
	(0.00860)	(0.0106)	(0.00935)	(0.01145)
Black	0.00531	0.0126**	0.00426	0.00544
	(0.00480)	(0.00592)	(0.00519)	(0.00633)
Hispanic	-0.00883	-0.00267	-0.00516	-0.00634
	(0.00705)	(0.00869)	(0.00771)	(0.00945)
Metro	-0.00118	0.00433	-0.0023	0.00232
	(0.00916)	(0.0113)	(0.00992)	(0.01215)

Notes: Each covariates are regressed on indicator for crossing cutoff, distance from cutoff, and interaction between indicator for crossing cutoff and distance from cutoff following equation (2). Here, "BW=45" and "BW=30" specification use observations within 45 and 30 days around the cutoff, respectively. Non-parametric estimation uses local polynomial RD point estimators with robust bias-corrected confidence intervals and inference procedures developed by Calonico et al. (2014)).

Table 3: Impact of Waiting on Standardized Scores: Alternative Specifications

	Math Score						
Model	All Grades	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
	Panel A: No Controls						
OLS	0.457***	0.479***	0.466***	0.445***	0.432***	0.490***	0.429***
	(0.0400)	(0.0339)	(0.0398)	(0.0367)	(0.0461)	(0.0630)	(0.0662)
2SLS	0.323***	0.341***	0.338***	0.355***	0.280***	0.318***	0.256***
	(0.0286)	(0.0352)	(0.0352)	(0.0406)	(0.0381)	(0.0515)	(0.0607)
	Panel B: Individual level controls						
2SLS	0.291***	0.327***	0.298***	0.305***	0.242***	0.302***	0.227***
	(0.0258)	(0.0341)	(0.0315)	(0.0384)	(0.0378)	(0.0486)	(0.0561)
	Panel C: Individual level controls + School FE						
2SLS	0.278***	0.326***	0.295***	0.289***	0.223***	0.266***	0.198***
	(0.0246)	(0.0337)	(0.0310)	(0.0364)	(0.0371)	(0.0427)	(0.0524)
Observations	187,747	45,184	39,279	33,468	27,822	23,508	18,486
	Reading Score						
Model	All Grades	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
	Panel A: No Controls						
OLS	0.452***	0.464***	0.462***	0.451***	0.438***	0.470***	0.428***
	(0.0387)	(0.0346)	(0.0365)	(0.0344)	(0.0430)	(0.0591)	(0.0598)
2SLS	0.350***	0.368***	0.361***	0.405***	0.328***	0.302***	0.273***
	(0.0258)	(0.0305)	(0.0309)	(0.0400)	(0.0434)	(0.0466)	(0.0545)
	Panel B: Individual level controls						
2SLS	0.319***	0.355***	0.322***	0.357***	0.290***	0.289***	0.244***
	(0.0252)	(0.0307)	(0.0312)	(0.0379)	(0.0442)	(0.0458)	(0.0512)
	Panel C: Individual level controls + School FE						
2SLS	0.314***	0.354***	0.329***	0.343***	0.276***	0.275***	0.237***
	(0.0238)	(0.0306)	(0.0292)	(0.0344)	(0.0463)	(0.0429)	(0.0480)
Observations	187,118	45,074	39,163	33,367	27,730	23,408	18,376

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: All treatment effects were estimated using 2SLS specification. Treatment is waiting an additional year to start kindergarten. Test scores are standardized by assessment system, year, and grade. Individual level controls are: female, FRL eligibility, white, black, and hispanic. 45 days as bandwidth has been used here for selecting observations close to the cutoff. Standard errors are clustered by school-districts.

Table 4: Impact of Waiting on Standardized Score: Heterogeneity Analysis

	Math Score		Reading Score	
	Sub-sample	Interaction	Sub-sample	Interaction
	Panel A: Female			
Waiting	0.371***	0.294***	0.394***	0.312***
	(0.0389)	(0.0359)	(0.0357)	(0.0333)
Waiting x Female		0.0563*		0.0774***
		(0.0303)		(0.0296)
Female		-0.0693***		0.133***
		(0.0239)		(0.0239)
Observationss	90,948	187,747	90,705	187,118
R-squared	0.031	0.031	0.034	0.041
	Panel B: FRL Eligible			
Waiting	0.280***	0.336***	0.321***	0.355***
	(0.0221)	(0.0406)	(0.0238)	(0.0368)
Waiting x FRL		-0.0642**		-0.0497*
		(0.0296)		(0.0284)
FRL		-0.627***		-0.631***
		(0.0615)		(0.0618)
Observations	85,774	187,747	85,395	187,118
R-squared	0.012	0.139	0.013	0.139
	Panel C: White			
Waiting	0.324***	0.271***	0.346***	0.311***
	(0.0504)	(0.0225)	(0.0476)	(0.0240)
Waiting x white		0.0781***		0.0410
		(0.0280)		(0.0315)
White		0.538***		0.515***
		(0.0649)		(0.0702)
Observations	129,074	184,155	128,706	183,540
R-squared	0.019	0.104	0.020	0.146
	Panel D: Metro			
Waiting	0.331***	0.324***	0.349***	0.358***
	(0.00951)	(0.0378)	(0.00562)	(0.0363)
Waiting x metro		-0.00323		-0.0162
		(0.0322)		(0.0305)
Metro		-0.0359		0.0191
		(0.174)		(0.164)
Observations	74,181	187,747	73,868	187,118
R-squared	0.035	0.032	0.038	0.033

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: All treatment effects were estimated using 2SLS specification considering 45 days of bandwidth around the cutoff. All the regressions are run without any controls. Standard errors are clustered at school-district level. Treatment is waiting an additional year to start kindergarten. Test scores are standardized by assessment system, year, and grade. Panel C regressions exclude Asian from the sample.

Table 5: Impact of Waiting on Standardized Score: Quantile Approach

Outcome	Conditional Quantile Treatment Effects				
variable	0.10	0.30	0.50	0.70	0.90
Panel A	Quantile Regression				
Math Score	0.347***	0.446***	0.492***	0.512***	0.527***
	(0.00824)	(0.00723)	(0.00740)	(0.00778)	(0.0113)
Constant	-1.441***	-0.833***	-0.346***	0.185***	1.012***
	(0.00800)	(0.00721)	(0.00749)	(0.00819)	(0.0121)
Reading Score	0.379***	0.480***	0.493***	0.488***	0.481***
	(0.00899)	(0.00785)	(0.00754)	(0.00768)	(0.0106)
Constant	-1.519***	-0.826***	-0.287***	0.250***	1.016***
	(0.00908)	(0.00810)	(0.00782)	(0.00822)	(0.0113)
Panel B	IVQR				
Math Score	0.209***	0.317***	0.330***	0.355***	0.358***
	(0.0180)	(0.0183)	(0.0209)	(0.0249)	(0.0409)
Constant	-1.347***	-0.761***	-0.254***	0.265***	1.074***
	(0.0134)	(0.0144)	(0.0170)	(0.0199)	(0.0318)
Reading Score	0.213***	0.351***	0.403***	0.441***	0.407***
	(0.0239)	(0.0230)	(0.0241)	(0.0288)	(0.0421)
Constant	-1.416***	-0.747***	-0.229***	0.282***	1.063***
	(0.0154)	(0.0156)	(0.0171)	(0.0205)	(0.0304)

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: Each column indicates a separate regression. All treatment effects were estimated using 2SLS specification considering 45 days of bandwidth around the cutoff. All the regressions are run without any controls. Treatment is waiting an additional year to start kindergarten. Panel A and B show results using OLS and 2SLS specifications respectively. IVQR estimates are obtained using smoothed estimating equations (SEE) outlined in Kaplan & Sun (2017). Test scores are standardized by assessment type, grade, and year. Robust standard error is used for inference purpose.

References

Angrist, J. D. & Krueger, A. B. (1992), 'The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples', *Journal of the American statistical Association* **87**(418), 328–336.

Attar, I. & Cohen-Zada, D. (2018), 'The effect of school entrance age on educational outcomes: Evidence using multiple cutoff dates and exact date of birth', *Journal of Economic Behavior & Organization* **153**, 38–57.

Autor, D., Figlio, D., Karbownik, K., Roth, J. & Wasserman, M. (2016), 'School quality and the gender gap in educational achievement', *American Economic Review* **106**(5), 289–295.

Barua, R. & Lang, K. (2016), 'School entry, educational attainment, and quarter of birth: a cautionary tale of a local average treatment effect', *Journal of Human Capital* **10**(3), 347–376.

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.

Bedard, K. & Dhuey, E. (2007), 'Is september better than january? the effect of minimum school entry age laws on adult earnings', *Unpublished manuscript* .

Calonico, S., Cattaneo, M. D. & Titiunik, R. (2014), 'Robust data-driven inference in the regression-discontinuity design', *The Stata Journal* **14**(4), 909–946.

Cannon, J. S. & Lipscomb, S. (2008), *Changing the kindergarten cutoff date: Effects on California students and schools*, Public Policy Institute of California San Francisco, CA.

Cascio, E. U. (2006), 'Public preschool and maternal labor supply: evidence from the introduction of kindergartens into american public schools'.

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.

Cattaneo, M. D., Jansson, M. & Ma, X. (2018), 'Manipulation testing based on density discontinuity', *The Stata Journal* **18**(1), 234–261.

Chernozhukov, V. & Hansen, C. (2004), 'The effects of 401 (k) participation on the wealth distribution: an instrumental quantile regression analysis', *Review of Economics and statistics* **86**(3), 735–751.

Chetty, R., Hendren, N., Lin, F., Majerovitz, J. & Scuderi, B. (2016), 'Childhood environment and gender gaps in adulthood', *American Economic Review* **106**(5), 282–288.

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.

Datar, A. (2006), 'Does delaying kindergarten entrance give children a head start?', *Economics of Education review* **25**(1), 43–62.

Depew, B. & Eren, O. (2016), 'Born on the wrong day? school entry age and juvenile crime', *Journal of Urban Economics* **96**, 73–90.

Elder, T. E. & Lubotsky, D. H. (2009), 'Kindergarten entrance age and children's achievement impacts of state policies, family background, and peers', *Journal of human Resources* **44**(3), 641–683.

Fertig, M. & Kluve, J. (2005), 'The effect of age at school entry on educational attainment in germany'.

Fletcher, J. & Kim, T. (2016), 'The effects of changes in kindergarten entry age policies on educational achievement', *Economics of education review* **50**, 45–62.

Gelbach, J. B. (2002), 'Public schooling for young children and maternal labor supply', *American Economic Review* **92**(1), 307–322.

Kaplan, D. M. & Sun, Y. (2017), 'Smoothed estimating equations for instrumental variables quantile regression', *Econometric Theory* **33**(1), 105–157.

Koenker, R. & Bassett Jr, G. (1978), 'Regression quantiles', *Econometrica: journal of the Econometric Society* pp. 33–50.

Lee, D. S. & Lemieux, T. (2010), 'Regression discontinuity designs in economics', *Journal of economic literature* **48**(2), 281–355.

Lenard, M. A. & Peña, P. A. (2018), 'Maturity and minorities: the impact of redshirting on achievement gaps', *Education Economics* **26**(6), 593–609.

Lincove, J. A. & Painter, G. (2006), 'Does the age that children start kindergarten matter? evidence of long-term educational and social outcomes', *Educational Evaluation and Policy Analysis* **28**(2), 153–179.

McCrary, J. & Royer, H. (2011), 'The effect of female education on fertility and infant health: evidence from school entry policies using exact date of birth', *American economic review* **101**(1), 158–195.

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.

Nollenberger, N., Rodríguez-Planas, N. & Sevilla, A. (2016), 'The math gender gap: The role of culture', *American Economic Review* **106**(5), 257–261.

Oshima, T. & Domaleski, C. S. (2006), 'Academic performance gap between summer-birthday and fall-birthday children in grades k-8', *The Journal of Educational Research* **99**(4), 212–217.

Peña, P. A. (2017), 'Creating winners and losers: Date of birth, relative age in school, and outcomes in childhood and adulthood', *Economics of Education Review* **56**, 152–176.

Puhani, P. A. & Weber, A. M. (2008), *Does the early bird catch the worm? Instrumental variable estimates of early educational effects of age of school entry in Germany*, Springer.

Ricks, M. D. (2022), 'Strategic selection around kindergarten recommendations'.

Shapiro, A. (2023), 'Over diagnosed or over looked? the effect of age at time of school entry on students receiving special education services', *Exceptional Children* **89**(2), 161–177.

Stipek, D. (2002), 'At what age should children enter kindergarten? a question for policy makers and parents. social policy report. volume 16, number 2.', *Society for Research in Child Development* .

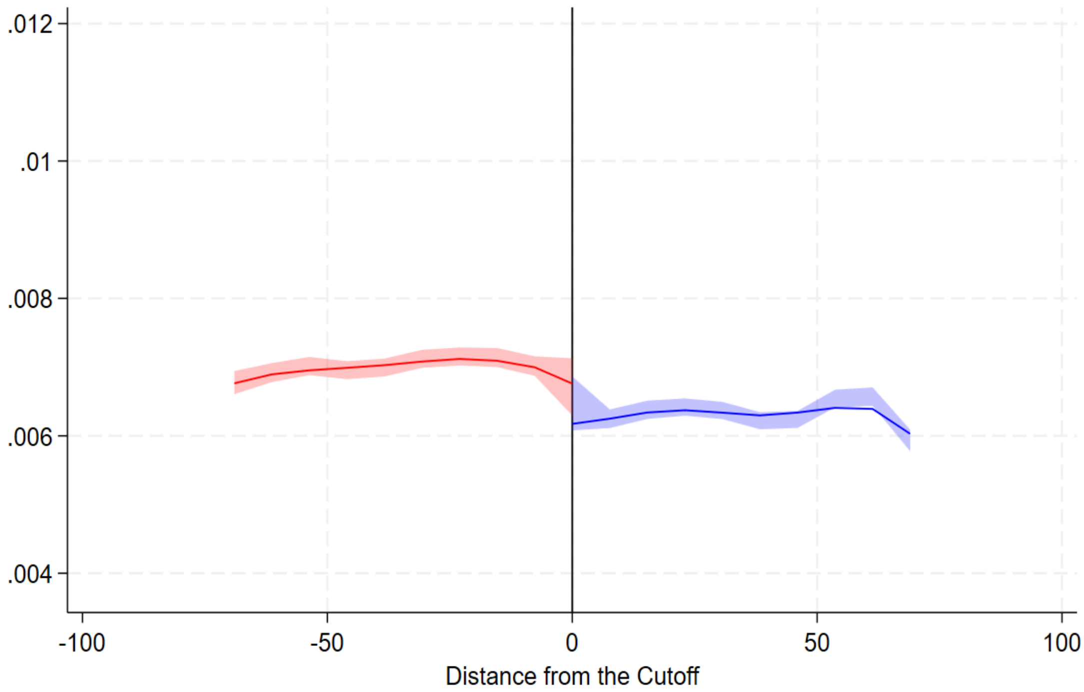
Taveras, E. (2021), 'An unintended effect of school entrance age: Pushing children ahead through private school'.

Thompson, A. H., Barnsley, R. H. & Battle, J. (2004), 'The relative age effect and the development of self-esteem', *Educational Research* **46**(3), 313–320.

Uphoff, J. K. & Gilmore, J. (1985), 'Pupil age at school entrance–how many are ready for success?.', *Educational leadership* **43**(1), 86–90.

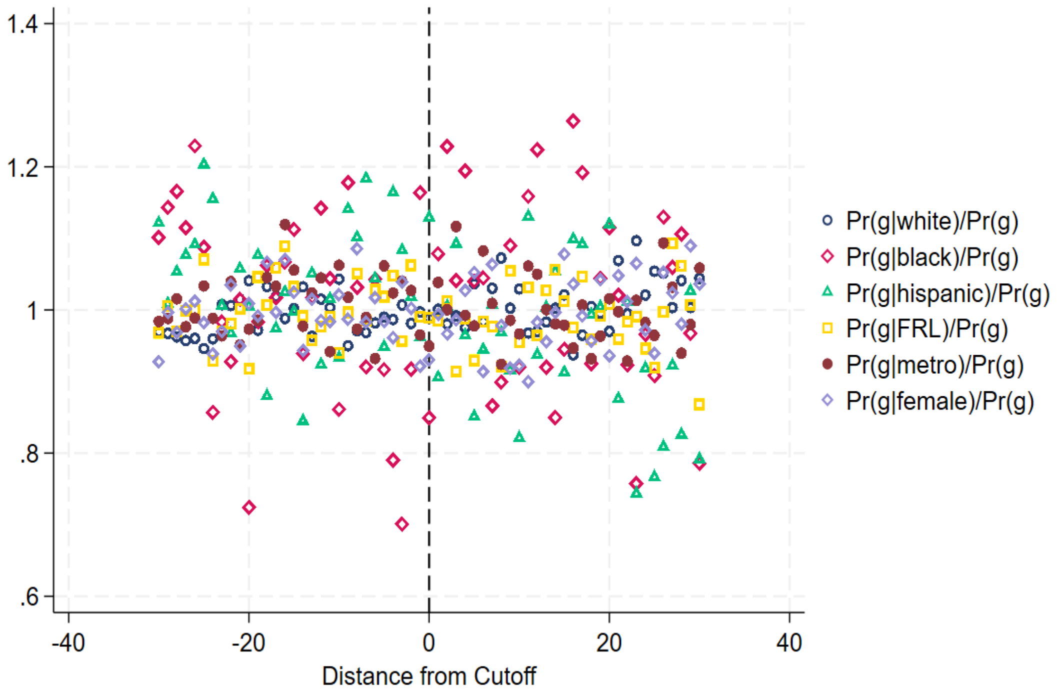
Zimmerman, S. D. (2014), 'The returns to college admission for academically marginal students', *Journal of Labor Economics* **32**(4), 711–754.

Figure A1: Manipulation Test



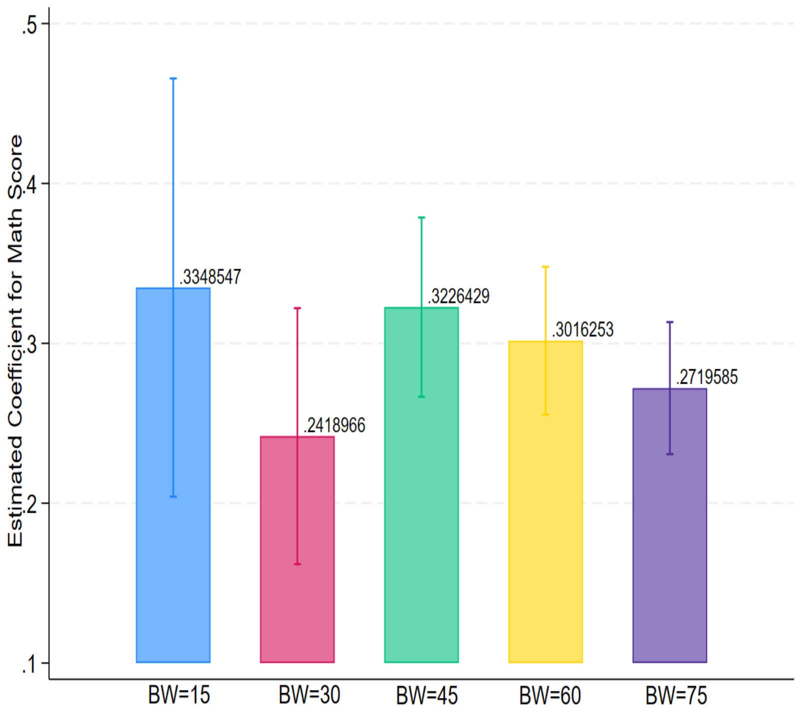
Notes: rddensity plot with local quadratic approximation producing T=0.3095 (p=0.7569).

Figure A2: Conditional Density Test

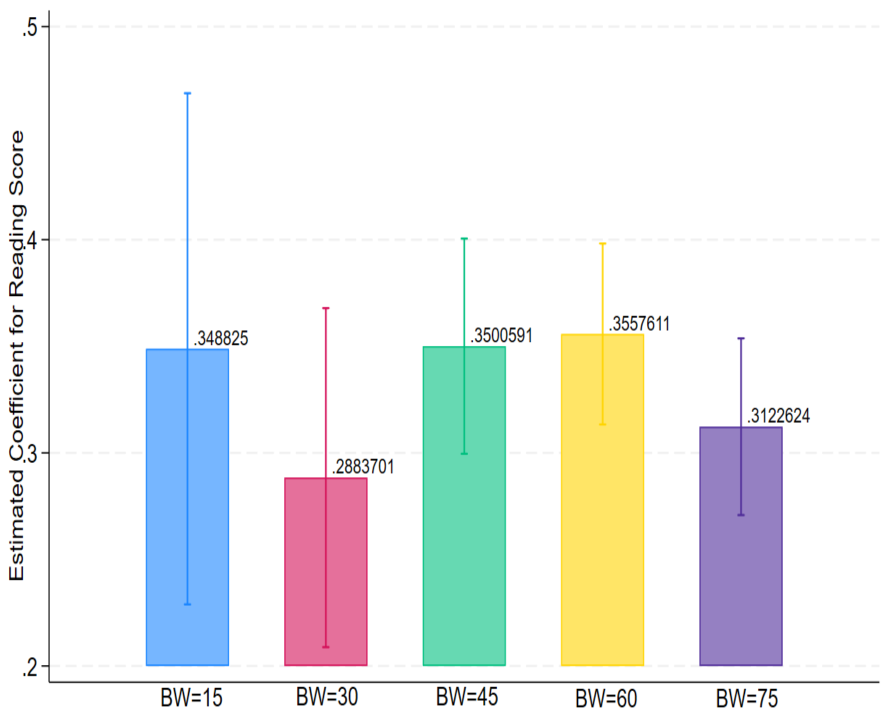


Notes: Ratios of conditional to unconditional densities of birthdate by the distance from the cutoff for six different groups: White, Black, Hispanic, female, FRL eligible students, and students living in metropolitan areas. Lincoln and Omaha are considered as metropolitan area. Densities are calculated with bin width of 1 day.

Figure A3: Checking Bandwidth Sensitivity



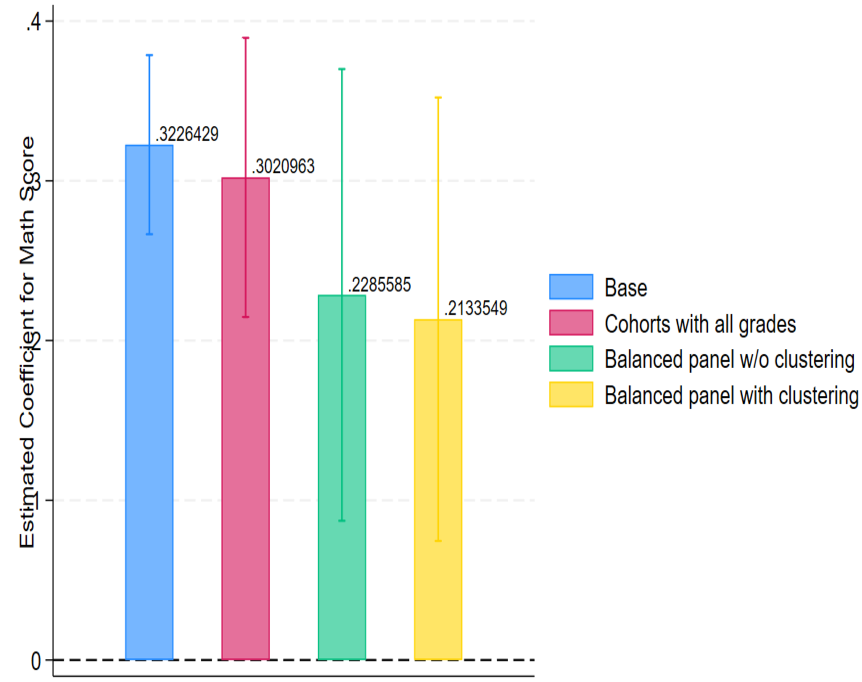
Bandwidth Sensitivity: Math score



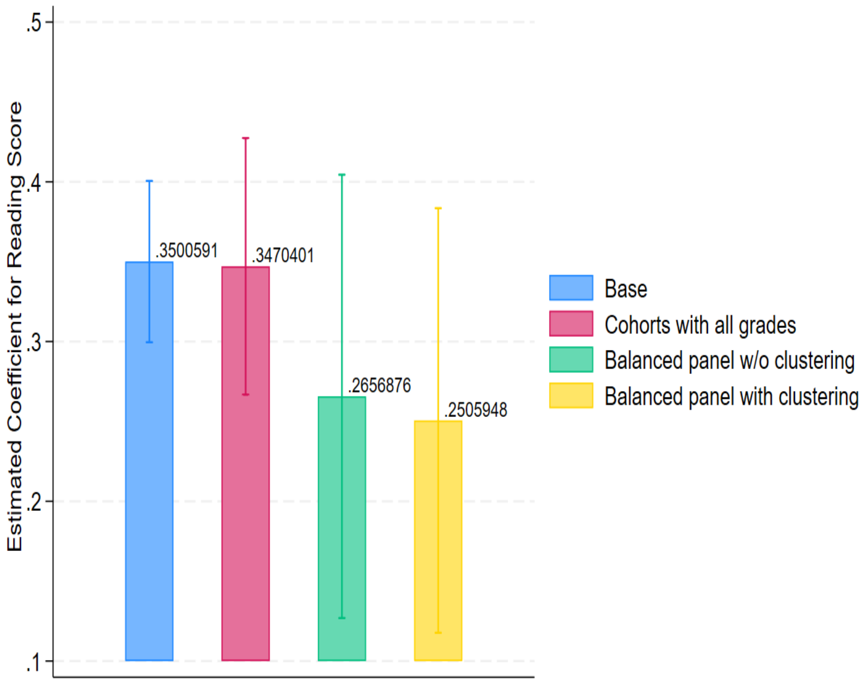
Bandwidth Sensitivity: Reading Score

Notes: Each bar indicates estimated coefficient from a separate regression with different bandwidth using 2SLS specification. Treatment is waiting an additional year to start kindergarten. These regressions do not include any controls. Test scores are standardized by assessment type, grade, and year. Standard errors are clustered at the school-district level.

Figure A4: Checking Sensitivity to Attrition



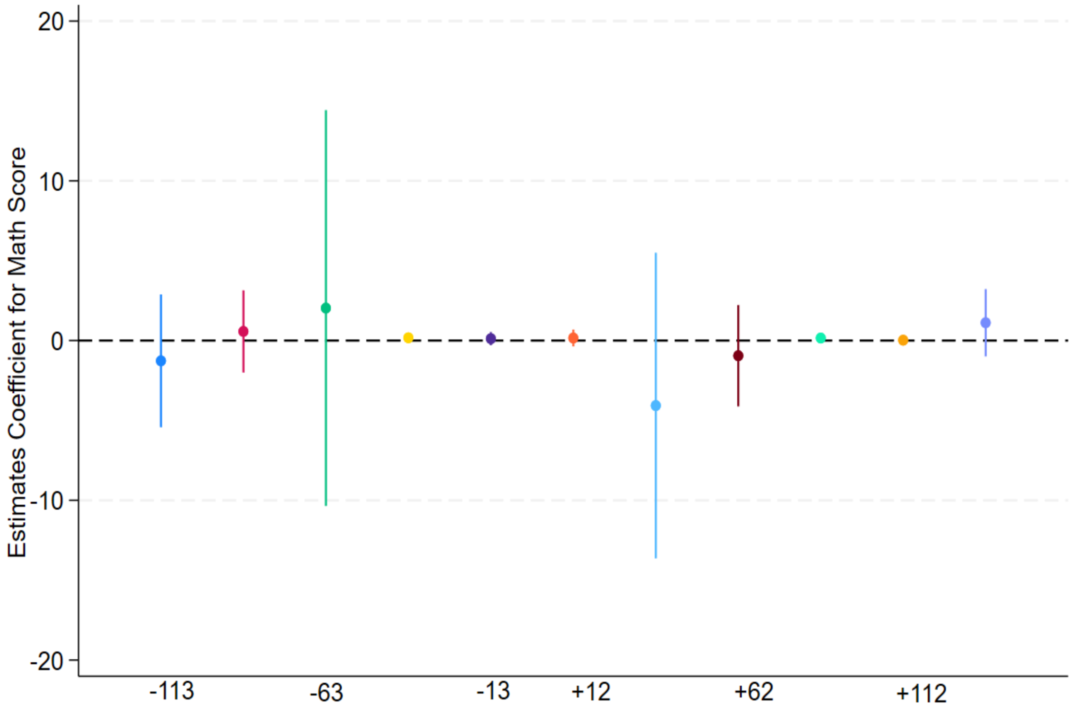
Sensitivity to Attrition: Math score



Sensitivity to Attrition: Reading score

Notes: Each bar indicates estimated coefficient from a separate regression considering 45 days of bandwidth around the cutoff using 2SLS specification. Each of them represents coefficient obtained from running regression with a different sample. The blue bar represents the base category. The pink bar shows regression results for cohorts with scores from all grades, particularly children who started kindergarten in 2007-08 and 2008-09. The green bar portrays results for a balanced panel without clustering, while the yellow bar incorporates clustering at the school district level. Treatment is waiting an additional year to start kindergarten. These regressions do not include any controls. Test scores are standardized by assessment type, grade, and year.

Figure A5: Checking Sensitivity to Different Cutoffs



Notes: The figure shows estimated coefficient for math score at pseudo cutoff around the original cutoff using 2SLS specification. All coefficients are estimated considering 45 days of bandwidth. Treatment is waiting an additional year to start kindergarten. These regressions do not include any controls. Test scores are standardized by assessment type, grade, and year. Standard errors are clustered at the school-district level.

Table A1: Descriptive Statistics

	Before change in cutoff		After change in cutoff	
	Before Oct cutoff	After Oct cutoff	Before Jul cutoff	After Jul cutoff
Waiting	0.18	1	0.09	0.95
	(0.38)	(0.05)	(0.29)	(0.22)
Gifted	0.17	0.20	0.11	0.14
	(0.38)	(0.40)	(0.31)	(0.34)
Special Educ	0.24	0.22	0.24	0.22
	(0.43)	(0.42)	(0.43)	(0.41)
Repetition	0.06	0.02	0.03	0.01
	(0.24)	(0.13)	(0.18)	(0.10)
Observations	84,662	21,245	51,065	34,428

Notes: Sample consists of public school students of Nebraska starting kindergarten from 2007-08 to 2015-16. All the students in the sample have started kindergarten at the age of 57 to 79 months. All the scores are standardized by assessment name, grade and year. Gifted, Special Educ, Repetition are coded as one if this identification occurs by grade 4. Total number of observations is 191,400.

Table A2: Summery Statistics by Quantiles

Panel A: Standardized Math Score							
Score Interval	Math score	Female	FRL eligible	White	Black	Hispanic	Metro
Q1	-1.224	0.484	0.694	0.504	0.139	0.271	0.454
	(0.404)	(0.50)	(0.461)	(0.5)	(0.346)	(0.445)	(0.50)
Q2	-0.375	0.508	0.515	0.664	0.063	0.206	0.366
	(0.188)	(0.50)	(0.50)	(0.472)	(0.242)	(0.405)	(0.482)
Q3	0.282	0.496	0.379	0.756	0.036	0.145	0.359
	(0.203)	(0.50)	(0.485)	(0.429)	(0.188)	(0.352)	(0.480)
Q4	1.319	0.453	0.241	0.826	0.020	0.088	0.396
	(0.546)	(0.498)	(0.428)	(0.379)	(0.140)	(0.283)	(0.489)
Panel B: Standardized Reading Score							
Score Interval	Reading score	Female	FRL eligible	White	Black	Hispanic	Metro
Q1	-1.280	0.428	0.687	0.524	0.125	0.267	0.429
	(0.465)	(0.495)	(0.464)	(0.499)	(0.331)	(0.442)	(0.495)
Q2	-0.332	0.478	0.519	0.659	0.066	0.208	0.365
	(0.201)	(0.50)	(0.50)	(0.474)	(0.248)	(0.406)	(0.482)
Q3	0.338	0.505	0.382	0.745	0.042	0.149	0.370
	(0.197)	(0.50)	(0.486)	(0.436)	(0.202)	(0.356)	(0.483)
Q4	1.276	0.532	0.239	0.823	0.024	0.086	0.409
	(0.458)	(0.499)	(0.426)	(0.382)	(0.153)	(0.280)	(0.492)

Notes: Test scores are standardized by assessment system, year, and grade. The proportion of individual level controls, such as, female, FRL eligibility, homelessness status, white, black, hispanic, and metro (Lincoln and Omaha are considered as metro) is calculated by dividing the sample into 4 quantiles based on math score (Panel A) & reading score (Panel B) and then calculating the average for all students within the quantile.

Table A3: Symmetry Around the Cutoff

Covariate	Before Cutoff	After Cutoff	P value of difference
Female	0.4838	0.4833	0.9011
FRL Eligible	0.4658	0.4634	0.6005
White	0.6864	0.6960	0.0249
Black	0.0706	0.0730	0.3112
Hispanic	0.180	0.1668	0.0002
Asian	0.0213	0.0221	0.5503
Metro	0.4110	0.4094	0.7389

Notes:1st column shows average of each covariates before the cutoff, while, 2nd column presents average after the cutoff. 45 days of bandwidth has been taken around the cutoff. 3rd column indicates the p-value of the difference between column 1 and column 2.

Table A4: Test for Constant Effect

Sample	Dependent Variable	K-S Statistic	Critical value	Result
All	Std math	7.931	2.507	Reject null
All	Std reading	7.220	2.462	Reject null

Notes: All the provided statistics are outcomes derived from the post-estimation command following SEE (Smoothed estimating equations) estimation. The null hypothesis is that the treatment effect is constant for all the estimated quantiles. Critical values are reported at 95 percent confidence level.

Table A5: Test Scores by Race

	White	Black	Hispanic	Asian
Math	0.1673	-0.652	-0.3832	0.4261
	(0.9761)	(0.8858)	(0.8817)	(1.009)
Reading	0.1599	-0.5724	-0.3826	0.3388
	(0.9719)	(0.9591)	(0.9056)	(1.017)

Notes: Average Math and Reading score for different demographic groups considering the full sample.

Table A6: Checking Attrition at Each Grade

	Missing Score					
Model	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
2SLS	0.00451	0.00541	0.00262	0.00431	-0.00916***	-0.00769***
	(0.00398)	(0.00378)	(0.00356)	(0.00331)	(0.00308)	(0.00277)

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Missing score is defined as one if an individual's score is missing for that particular grade. Missing score for each grade are regressed on crossing cutoff, distance from cutoff, and interaction between crossing cutoff and distance from cutoff following equation (2). 45 days as bandwidth has been used here for selecting observations close to the cutoff. These regressions do not include any controls.