Stanford cepa | Center for Education Policy Analysis

CEPA Working Paper No. 15-08

The Gift of Time? School Starting Age and Mental Health

AUTHORS

Thomas Dee
Stanford University

Hans Henrik Sievertsen

The Danish National Centre for Social Research

ABSTRACT

In many developed countries, children now begin their formal schooling at an older age. However, a growing body of empirical studies provides little evidence that such schooling delays improve educational and economic outcomes. This study presents new evidence on whether school starting age influences student outcomes by relying on linked Danish survey and register data that include several distinct, widely used, and validated measures of mental health that are reported out-of-school among similarly aged children. We estimate the causal effects of delayed school enrollment using a "fuzzy" regression-discontinuity design based on exact dates of birth and the fact that, in Denmark, children typically enroll in school during the calendar year in which they turn six. We find that a one-year delay in the start of school dramatically reduces inattention/hyperactivity at age 7 (effect size = -0.7), a measure of self regulation with strong negative links to student achievement. We also find that this large and targeted effect persists at age 11. However, the estimated effects of school starting age on other mental-health constructs, which have weaker links to subsequent student achievement, are smaller and less persistent.

VERSION

October 2015

Suggested citation: Dee, T., & Sievertsen, H.H. (2015). The Gift of Time? School Starting Age and Mental Health (CEPA Working Paper No.15-08). Retrieved from Stanford Center for Education Policy Analysis: http://cepa.stanford.edu/wp15-08

The Gift of Time? School Starting Age and Mental Health
Thomas S. Dee and Hans Henrik Sievertsen
NBER Working Paper No. 21610
October 2015
JEL No. I1,I2

ABSTRACT

In many developed countries, children now begin their formal schooling at an older age. However, a growing body of empirical studies provides little evidence that such schooling delays improve educational and economic outcomes. This study presents new evidence on whether school starting age influences student outcomes by relying on linked Danish survey and register data that include several distinct, widely used, and validated measures of mental health that are reported out-of-school among similarly aged children. We estimate the causal effects of delayed school enrollment using a "fuzzy" regression-discontinuity design based on exact dates of birth and the fact that, in Denmark, children typically enroll in school during the calendar year in which they turn six. We find that a one-year delay in the start of school dramatically reduces inattention/hyperactivity at age 7 (effect size = -0.7), a measure of self regulation with strong negative links to student achievement. We also find that this large and targeted effect persists at age 11. However, the estimated effects of school starting age on other mental-health constructs, which have weaker links to subsequent student achievement, are smaller and less persistent.

Thomas S. Dee
Stanford University
520 Galvez Mall, CERAS Building, 5th Floor
Stanford, CA 94305-3084
and NBER
tdee@stanford.edu

Hans Henrik Sievertsen
The Danish National Centre for Social Research
Copenhagen, Denmark
hhs@sfi.dk

1 Introduction

Over the last half century, the age at which children in the United States initiate their formal schooling has slowly increased. Historically, U.S. children attended kindergarten as five-year olds and first grade as six-year olds. However, roughly 20 percent of kindergarten students are now six years old (e.g., The New York Times, 2010; The Boston Globe, 2014). This "lengthening of childhood" reflects in part changes in state laws that moved forward the cutoff birth date at which 5 year olds were eligible for entering kindergarten (Deming and Dynarski, 2008). However, most of the increase in school starting ages is due to academic "redshirting"; an increasingly common decision by parents to seek developmental advantages for their children by delaying their school entry (i.e., the "gift of time"). Redshirting is particularly common for boys and in socioeconomically advantaged families (Bassok and Reardon, 2013).[1] Delayed school starts are also common in other developed countries. For example, in Denmark, one out of five boys and one out of ten girls have a delayed school start.[2] The conjectured benefits of starting formal schooling at an older age reflect two broad mechanisms. One is relative maturity; students may benefit when they start school at an older age simply because they have, on average, a variety of developmental advantages relative to their classroom peers. The second mechanism, absolute maturity, reflects the hypothesis that formal schooling is more developmentally appropriate for older children. Specifically, a literature in developmental psychology suggests that children who start school at a later age benefit from an extended period of informal, play-based preschool that complements language development and the capacity for "self regulation" of cognitive and emotional states (Vygotsky, 1978; Whitebread, 2011).

The decision of whether to delay a child's formal schooling is a recurring topic in the popular press (e.g., The New Yorker, 2013) with most coverage suggesting that there are educational and economic benefits to delayed school entry. However, the available research evidence largely suggests otherwise. A number of early studies (e.g., Bedard and Dhuey, 2006) did indeed show that children who start school later have, on average, higher per-

[1]For example, according to the U.S. National Center for Education Statistics 14% of the children who delayed school entrance in 2010 were children of parents in the lowest quintile of socioeconomic status, while 24% were children of parents in the highest quintile. The measure of socioeconomic status is based on parental education, occupation, and household income at the time of data collection.

[2]See Appendix Figure A.1 for the development of red-shirting in Denmark. Throughout this paper we refer to school starting age as the age at which a child enters kindergarten, which in Danish is called grade zero or "Børnehaveklasse".

formance on in-school tests (i.e., even after adjusting for the endogenous decision to redshirt). However, more recent studies suggest that these findings simply reflect the fact that children who start school later are older when the test is given.[3] For example, using Norweigan data, Black, Devereux, and Salvanes (2011) find that a higher school starting age implies a small, *negative* effect on an IQ test taken outside of school at age 18. In a related regression-discontinuity study using Swedish data, Fredriksson and Öckert (2013) conclude that a higher school starting age increased educational attainment slightly but not lifetime earnings.

In this study, we examine the causal effect of higher school starting age on different dimensions of mental health among similarly aged Danish children. There is some limited evidence that delays in school starting age improve measures of children's mental health (e.g., Black, Devereux, and Salvanes, 2011; Mühlenweg, Blomeyer, Stichnoth, and Laucht, 2012).[4] Our study contributes to this literature in several ways. First, we base our study on a unique source of data, a recent and large-scale survey of Danish children (the Danish National Birth Cohort or DNBC). The DNBC includes, for children at age 7 and at age 11, data from a widely used and validated mental-health screening tool (i.e., the Strengths and Difficulties Questionnaire or SDQ). The SDQ was was explicitly designed for children and generates measures of several distinct psychopathological constructs, some of which are clearly related to the theorized effects of delayed school starts. Second, we are able to credibly identify the effects of a delayed school start through a "fuzzy" regression discontinuity design based on *exact* day of birth. We identified the day of birth and school starting age of children in the DNBC by matching these data to population data available in the Danish administrative registry and Ministry of Education records. In Denmark, children are supposed to enter school in the calendar year in which they turn six. Using data on children's exact date of birth, we find that school starting age does indeed "jump" discontinuously for children born January 1st or later relative those born December 31st or earlier. We also avoid confounds due to "age at test" because the DNBC data are based on children of roughly the same age.[5] Finally, the Danish context (i.e., a universal day-care system with a centrally

[3]Angrist and Pischke (2008) offer this as an example of a "fundamentally unidentified" research question. A student's school starting age equals by definition their current age minus their time in school. So, for measures of in-school performance, the effects of school starting age cannot be disentangled from age-at-test and time-in-school effects.

[4]Elder and Lubotsky (2009) also find that students who delay formal schooling are less likely to receive a diagonsis of Attention Deficit/Hyperactivity Disorder (ADHD) but suggest that this is a diagnostic response to the *preschool* differences of children who start late.

[5]However, this does imply a potential collinearity with years of formal schooling. Children who start

specified structure) implies a fairly homogenous condition for students with higher school starting ages.[6]

Our results indicate that a one-year increase in the school starting age leads to significantly improved mental health (i.e., reducing the “difficulties” scores at age 7 by 0.6 SD. Interestingly, we find that these effects are largely driven by a large reduction (effect size = -0.70) in a single SDQ construct: the SDQ’s inattention/hyperactivity score (i.e., a measure of self regulation). Consistent with a literature that emphasizes the importance of self-regulation for student outcomes, we find that this construct is most strongly correlated with the in-school performance of Danish children. This targeted effect is also consistent with theoretical explanations from developmental psychology that stress the salience of extended play for the development of self regulation. We are also able to examine whether these short-term effects persist using the most recently available data which tracks students to age 11. We find that the large and concentrated effects largely persist to later childhood (i.e., an effect size for inattention/hyperactivity of -0.68). However, we also find evidence that these effects are heterogeneous. Using an approach recently introduced by Bertanha and Imbens (2014), we present evidence on the heterogeneity that distinguishes the "compliers" from the "never takers" and "always takers" in our "intent to treat" (ITT) design. We also find evidence that these effects are most pronounced among children with higher-earning, better-educated parents.

This paper proceeds as follows: Section Two provides brief discussions of the theoretical relationships between school-starting delays and child outcomes and of the existing empirical literature on this topic. Section Three introduces this study’s data, particularly the DNBC and the SDQ measures. Section Four discusses the Danish setting, presents evidence on the discontinuous jump in school starting age for children whose birthday are past the cutoff. This section also introduces our RD design and presents initial evidence on its validity. Section Five presents the results. Section Six concludes this paper.

kindergarten late necessarily have fewer years of formal school when their outcomes are observed. We argue that our pattern of results suggests effects related to a delayed school start. We also engage the question of whether our results may be due to reference biases in the rating of children (Elder, 2010).

[6]According to Statistics Denmark more than 95 percent of pre-school children are in daycare. In the US, in contrast, 27 percent of the delayed school entrants in 2010 were not in a non-parental arrangement according to the US National Center for Education Statistics.

2 Theoretical Framework and Prior Literature

One rationale for the growing number of parents who choose to delay their children's school starting age involves the perceived benefits of relative maturity for young children. This conjecture, recently popularized by Malcolm Gladwell's 2008 book, *Outliers*, turns on the claim that children who are slightly older than their peers experience early successes that are then followed by recursive processes of reinforcement and support.[7] A second class of rationales for delayed school starting age turns on the perceived benefits of increasing the absolute maturity of children when they begin formal schooling. That is, a delay in formal schooling may benefit student outcomes because slightly older children are more developmentally aligned with the demands and opportunities of formal schooling.

Recent economic models of skill formation (e.g., Cunha, Heckman, Lochner, and Masterov, 2006) emphasize the relevance of such dynamics (i.e., the relevance of prior skill for future skill formation). The importance of this alignment is also a longstanding theme in developmental psychology that draws on the influential work of Vygotsky (1978). Specifically, this literature suggests that engagement with teachers and student peers is uniquely effective when the skills to be learned fall within a student's "zone of proximal development" (i.e., they cannot be mastered individually but can with the guidance of a knowledgeable person). Students who start school later and with more maturity may be more likely to experience this productive instructional alignment. Another related and theorized benefit to delayed school starting ages involves the conjectured importance of play in child development. Pretend play is thought to promote abstract and symbolic reasoning because it demands the separation of actions and objects from reality (e.g., a banana as a phone). Furthermore, pretend play necessitates intellectual and emotional self-regulation in that children in play will sustain and bound their fictional realities. The developmental relevance of play is reflected in the oft-quoted line: "in play, it is as though [the child] were a head taller than himself" (Vygotsky, 1978, p.102). Given that children who delay their school starting age are typically in home care or less formal preschool, they may benefit from an extended experience in relatively playful environments.

[7] Though parents' belief in the gains from relative maturity may be widespread, the empirical evidence on the direct educational benefits from a higher relative age is at best equivocal. In particular, a random-assignment study by (Cascio and Schanzenbach, forthcoming) finds that students who are old for their cohort may have poorer outcomes because of peer-group effects. To the extent that such effects exist in our Danish data, it implies that we are understating the targeted mental-health benefits of a higher school starting age.

In recent years, several empirical studies have attempted to identify the reduced-form effect of a higher school starting age (i.e., inclusive of relative and absolute age mechanisms) on near-term child development and longer-run outcomes. For example, using data from a sample of 20 countries, Bedard and Dhuey (2006) find that being older at school enrollment generally increases fourth and eight grade test scores in mathematics and science. The identification strategy in this study relied on a student's predicted school starting age (i.e., using their birth date and country-specific cutoff dates) as an instrumental variable for their actual school starting age. Using this basic approach, several other country-specific studies find that students who start school later score substantially higher on in-school tests (e.g., Puhani and Weber, 2005; Crawford, Dearden, and Meghir, 2007; McEwan and Shapiro, 2008). However, a more recent literature has suggested that these estimated effects of school starting age on in-school test performance are overstated because of the collinearity between a student's age at test and their school starting age. That is, students with higher school starting ages may perform better on in-school tests simply because they are older than those who started earlier.[8]

For example, Elder and Lubotsky (2009) show that older children perform substantially better in the fall of their kindergarten year *before* the onset of formal schooling could have had much effect. Furthermore, there is evidence that the effects of school starting age decline over time (Elder and Lubotsky, 2009; Cascio and Schanzenbach, forthcoming). An influential study by Black, Devereux, and Salvanes (2011) separately identifies the effects of school starting age and age at test, using unique institutional circumstances and data from Norway. In particular, they rely on IQ tests taken by males at 18 as part of their mandatory military service. They identify the effects of school starting age and age at test on these IQ scores, using each person's birth month and the cutoff dates for entering school and the age at which the IQ test is taken (i.e., akin to a regression-discontinuity design). They find that age at test has large positive effects on measured IQ while school starting age has quite small but statistically significant *negative* effects. They also find that school starting age has no effect on educational attainment nor on long-run earnings. In a related study, Fredriksson and Öckert (2013) use Swedish data on birth cohorts from 1935 to 1955 in a regression-discontinuity framework and show that being older at school enrollment

[8]Aliprantis (2014) also argues that the instrumental variable used in these application may suffer from monotonicity violations and states that "the best evidence on the effects of redshirting is likely to be found in studies employing regression discontinuity designs."

increases educational attainment. Interestingly, their sample period spans the introduction of a school reform that postponed tracking. They find that, when tracking was delayed, the effects of a child's school starting age are smaller. While they find that the effects on discounted life-time earnings on average is very small to negative, they also find positive earnings effects of school starting age for individuals with low-educated parents.

Studies that have examined the effects of school starting age on non-schooling and non-market outcomes produce similary equivocal evidence. For example, Black, Devereux, and Salvanes (2011) find that a delayed school starting age increases the likelihood that a girl will give birth within 12 years of starting her formal schooling (i.e., interrupting her human-capital accumulation). A regression-discontinuity study based on exact date of birth and Danish data (Landersø, Nielsen, and Simonsen, 2013) finds that an increased school starting age reduces the propensity to commit crime. However, this result appears to be driven by incapacitation rather than a developmental effect.

In the context of this study, the limited, prior evidence that has focused on dimensions of mental health has particular relevance. For example, Elder and Lubotsky (2009) find that children with delayed school starts are less likely to be diagnosed with Attention Deficit/Hyperactivity Disorder (ADHD) between kindergarten and fifth grade. However, they suggest that this effect reflects parents and schools responding to the *prekindergarten* differences among students with differences in school starting ages, which fade at older ages. A recent, small-scale study of 360 children from the Rhine-Neckar region in central Germany (Mühlenweg, Blomeyer, Stichnoth, and Laucht, 2012) finds that later school starting ages imply more persistence and less hyperactivity at age 8 relative to the levels observed prior to school entry. Finally, the study of Norwegian military records by Black, Devereux, and Salvanes (2011) find that a year of delayed school entry reduces the chance that males are reported as having mental health problems (i.e., half a percentage point where the mean rate of poor mental health is 7 percent). This measure is based on a psychologist's assessment from an interview to determine each young man's suitability for military service.

In sum, this body of theoretical hypotheses and empirical evidence suggests a puzzling contrast. Parents and policymakers are increasingly choosing to delay childrens' school starting ages. And there are several theoretical reasons to suspect that these delays confer developmental advantages (e.g., relative age effects, the dynamic complementarity in skill formation, extending pre-school periods of play). Yet there is not strong evidence that a delayed school start meaningfully improves key educational and economic outcomes. In contrast, there is some limited suggestive evidence that a higher school starting may improve measures of psychological adaptation and mental health. This study presents new evidence on the effects of school starting age on dimensions of mental health. This new evidence advances this literature in several ways. First, we are able to rely on data from a widely used and extensively validated mental health screening tool that is specifically designed for children and teens and that measures several diagnostically relevant constructs. Second, because these data are collected among children at the same ages (i.e., age 7 and again at age 11), we avoid confounds related to "age at test." Third, we are able to match these unique data to the students' exact dates of birth and to implement a regression-discontinuity design that provides credibly causal evidence on the effects of a delayed school start. Finally, we also argue that our pattern of results speaks indirectly to the empirical salience of absolute and relative-age mechanisms. In the next two sections, we describe these data and methods in more detail before turning to our results.

3 Data

We create our analysis samples by matching children included in the Danish National Birth Cohort Survey (DNBC, Olsen, Melbye, Olsen, Sørensen, Aaby, Andersen, Taxbøl, Hansen, Juhl, Schow, et al., 2001) to data available for the full Danish population from the national administrative registers. The DNBC provides detailed measures of children's mental health at ages 7 and 11. The national administrative registers provide information on the child's birthday (i.e., the forcing variable in our regression-discontinuity design) as well as data on child and family traits at baseline. We describe each of these data sets in more detail below.

3.1 The Danish National Birth Cohort (DNBC)

The DNBC is a Danish nation-wide cohort study based on a large sample of women who were pregnant between 1996 and 2002 (i.e., roughly 10 percent of the births in the population during this period). Nearly 93,000 woman participated in the baseline interviews (i.e., during pregnancy). The fifth wave of the survey was fielded when the sampled child was approximately 7 years old. And, in 2014, a sixth survey wave that elicited information when the sampled children were age 11, concluded.[9] These surveys included questions on a diverse set of behaviors and traits such as risky behaviors (e.g., smoking and drinking), employment, and the health status of the child and the mother. During the fifth survey wave, the respondent was also asked to identify when the child *first* started kindergarten, which we use to identify their school starting age.[10] Critically, the fifth and sixth survey waves also included the 25-item Strengths and Difficulties Questionnaire (SDQ), which we describe in more detail below. It should be noted that the response rate to the DNBC does fall with the follow-up surveys. For example, 57,280 mothers participated in the fifth interview (child aged seven years), with 54,251 providing valid data on the SDQ and school starting age. Nearly 36,000 of these respondents also participated in the sixth survey wave (child aged 11 years). This survey attrition appears to be non-random and implies an external-validity caveat to our study. We provide data comparing the survey population and the full population in the Appendix Table A.1. In brief, mothers in the follow-up samples are, on average, more affluent (i.e., mothers with higher income, more schooling) and their children had higher birth weights. These mothers are also more likely to be married and in the labor force (Jacobsen, Nohr, and Frydenberg, 2010). However, we find that participation in these surveys is balanced around the birthday threshold so this non-response does not appear to threaten the internal validity of our RD design.

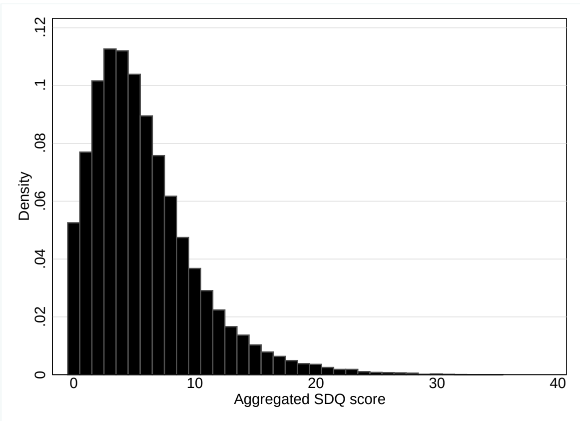
3.2 The Strengths and Difficulties Questionnaire (SDQ)

The SDQ is a mental-health screening tool designed specifically for children and teens and is in wide use internationally both in clinical settings and in research on child development. The questionnaire, which was developed by English child psychiatrist Robert N. Goodman in the mid 1990s, consists of 25 items (Goodman, 1997) that may describe the child in question. Examples of the items include "Restless, overactive, cannot stay still for long" and "Good attention span, sees work through to the end." For each item, the rater is asked to "consider the last 6 months" and to mark the description of the child in one of three ways: Not True, Somewhat True, Certainly True. The established scoring procedure for the

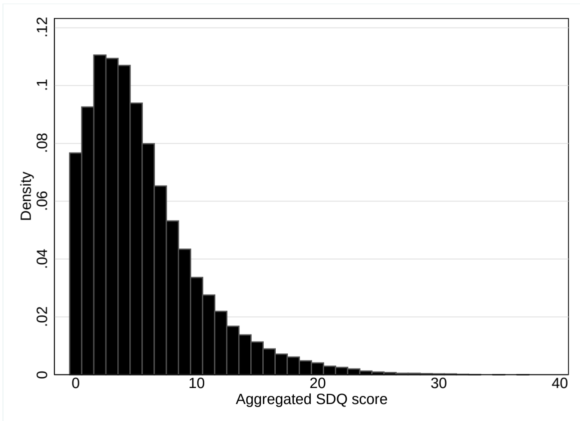
[9]Each survey wave was fielded on a rolling basis so as to get child data at roughly the same age. Differential response times necessarily create some variation in the age at observation. However, we control for each child's age at the time of interview and find that this age is well balanced around the threshold in our RD design.

[10]We find similar results when we use school starting age imputed from the subsequent National Test data we describe below. However, these parent reports are our preferred source of data on school starting age because repetition of kindergarten is not uncommon (i.e., roughly 10 percent), complicating the imputation of school starting age based on birthdates and when grade-level tests are taken

SDQ links each of the 25 items to one (and only one) of five distinct subscores: emotional symptoms, conduct problems, inattention/hyperactivity, peer problems, and a pro-social scale. Each subscore has five uniquely linked items and the response to each item is scored as 0, 1, or 2. The value for the subscore is simply the sum of the ratings for its five linked items. So, each subscore has a range of 0 to 10. The total "difficulties" score is the sum of the subscales, excluding the pro-social score, and can range from 0 to 40. For this difficulties score, values between 0 and 13 are regarded as normal, while scores 14-16 are borderline and scores from 17 to 40 are regarded as abnormal. For the pro-social scale 6-10 is normal, 5 is borderline, and 0-4 is abnormal. In our main analyses, we standardize each score (i.e., using the full population in each survey wave) so that our coefficients of interest can be interpreted as effect sizes. However, we also present linear probability models for the probability of an abnormal rating. Figure 1 shows the distribution of SDQ scores in our DNBC samples.



(a) *Age 7 Wave*



(b) *Age 11 Wave*

Figure 1: *The SDQ Total Difficulties Score*

The development of the SDQ items (and their scaling) was conducted with reference to the main categories of child mental-health disorders recognized by contemporary classification systems like the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (American Psychiatric Association, 1994). Psychometric studies have generally confirmed the convergent and discriminant validity of the five-factor structure of the SDQ in a variety of populations (Achenbach, Becker, Döpfner, Heiervang, Roessner, Steinhausen, and Rothenberger, 2008), though some studies suggest there should be fewer subscores.[11] Furthermore,

[11]The standard aggregation procedure is described on the website, www.sdqinfo.com. We independently examined the item-level responses in our DNBC data using a principal component analysis (PCA). The PCA revealed the same five dimensions as the standardized procedure.

in *both* the parent and teacher versions, the SDQ has demonstrated satisfactory internal consistency, test-retest reliability, and inter-rater agreement (e.g., Achenbach, Becker, Döpfner, Heiervang, Roessner, Steinhausen, and Rothenberger, 2008; Stone, Otten, Engels, Vermulst, and Janssens, 2010). The SDQ produces scores that are highly correlated with those from earlier prominent screening devices, the Rutter questionnaire and the Child Behavior Checklist (Goodman, 1997; Goodman and Scott, 1999). However, the SDQ also appears to have some conceptual and practical advantages. For example, the SDQ includes items related to strengths rather than just difficulties. The SDQ also has a strong comparative advantage in identifying constructs of contemporary relevance such as inattentiveness and sociability. The SDQ also has a practical advantage is that its 25-item format is substantially shorter (e.g., the school-age CBCL has 120 items).

To understand the properties of the SDQ subscores in our particular research context, we also examined how the SDQ scores of children in the DNBC predicted their in-school test performance on the Danish National Tests in two subjects (Danish and mathematics). These tests are obligatory annual assessment of children's cognitive skills in grades two through eight that are used as a tool for the teacher to assess the child's development. Specifically, we separately regressed the performance in three tests in Danish and two tests in mathematics on the five SDQ subscores measured at age 7. In each regression we include school fixed effects so that we are effectively making comparisons among students in the same schools. We also control for age at test, both for the SDQ scores and the in-school tests, by means of monthly indicators. Our results indicate that the peer-problems subscore is unrelated to future test performance. There are also somewhat anomalous results. Prosocial scores predict lower test scores in both subjects and all grades (i.e., effect sizes of 0.04 and 0.05). And emotional symptoms predict higher performance in Danish (effect size = 0.03). However, our main finding is that the two constructs associated with "externalizing behavior" - the conduct and inattention/hyperactivity constructs - strongly predict lower test performance across all grades and subjects. The effect sizes for conduct problems range from 0.05 to 0.07. And a 1 SD increase in the inattention/hyperactivity score predicts a reduction in future test performance ranging from 0.14 SD to 0.16 SD.

The uniquely strong link between the inattention/hyperactivity subscore and future student performance is noteworthy but not necessarily surprising.The inattention/hyperactivity construct is effectively synonymous with the concept of self regulation (i.e., the voluntary control of impulses in service of desired goals; Blake, Piovesan, Montinari, Warneken, and Gino (2014)). And an extensive literature has documented the importance of such self-regulation for student success (e.g., Duckworth and Carlson, 2013).[12] Interestingly, one of the theorized mechanisms through which higher school starting ages are thought to be developmentally beneficial, involves self-regulation. In particular, the extended periods of pretend play available to children who delay their school start may enhance their capacity for this important psychological adaption.

3.3 The Danish administrative registers

The Danish administrative data actually consists of several individual registers including the birth records, the income registers, and the education registers. All datasets are hosted by Statistics Denmark and linked by a unique personal identifier. The critical variable we draw from the registers forms the basis for the forcing variable in our RD design (i.e., the exact date of birth). However, we also use the registers to construct a variety of other family and child-specific control variables. For the children, we use information from the registers on birth weight, 5 minute APGAR score, and gestational age.[13] For the parents we use information on gross annual income, educational attainment, civil status, origin and age. We also record the number of siblings (living in the household) when the child is two years old using register data. Before we link the children to their parents and siblings we adjust the birth year to run from July to June instead of January to December. For example all children born in the period July 2000 to June 2001 are merged to parents' characteristics for the calendar year January to December 1999.

In Table 1, we show descriptive statistics for the key variables from our linked DNBC and register data, separately for both the age-7 and age-11 samples. These variables include our standardized SDQ measures, school starting age, a binary indicator for a delayed school starting age (i.e., a school starting age greater than 6 years, 7 months), the birth date centered on the January 1st threshold, a binary indicator for a birth date between January 1st and June 30 (i.e., our "intent to treat" measure), and a variety of baseline covariates. We standardized our SDQ measures using the mean and standard deviation specific to each

[12]The concept of self-regulation is also widely thought to be equivalent to the "Big 5" construct of conscientiousness, another highly outcome-relevant personality trait. Heckman and Kautz (2012) note "conscientiousness – the tendency to be organized, responsible, and hardworking—is the most widely predictive of the commonly used personality measures."

[13]The APGAR score is an evaluation of the infants' health measured on a 0-10 scale (where 10 is the best).

measure and the full population in each survey wave. A total of 54,251 parents completed both the SDQ and school starting age questions in the fifth wave of the survey (i.e., in years 2004-2010 when the sampled child was 7 years old). During the sixth survey wave (i.e., 2010-2014 when the sampled child was roughly 11 years old) 35,902 of these respondents participated in the DNBC.

4 Empirical framework

4.1 The Danish Context

Before they begin formal schooling, most children in Denmark (i.e., over 95 percent) are in daycare that is publicly provided and organized at the municipal level. Child care consists of center-based nurseries and family day care for children aged 1 to 3 years and daycare for children aged 3 to 6. In addition to the center-based nurseries, municipalities also fund family day care. The standards required of center-based day care and their staff are high compared to other OECD countries (Datta Gupta and Simonsen, 2010). For example, there is a high staff-child ratio and all permanent day care staff must have a pedagogical education. The requirements of family day care are lower.

Compulsory schooling begins in "grade zero" (also called kindergarten class) in August of the year in which the child turns six. Until 2009 grade zero was not mandatory, but 98% of children attended anyway (Browning and Heinesen, 2007). Compulsory schooling ends after ten years of schooling or in August of the year the child turns 17. Figure 2 summarizes the timing of events in childhood. The children typically do not change institution or class after they enrolled in grade zero (i.e., most children stay in the same class within the same school from grade zero until grade nine). After leaving compulsory education, the individual can choose between three-year upper secondary school (high school), vocational training (apprenticeship), or the labor market. Completing high school also allows access to higher education.

As children are supposed to enroll in school the year they turn six, school starting age should jump discontinuously as birthdays change from December 31 to January 1. To illustrate this institutional feature, we compare the events in Figure 1 for a child born December 31 to a child born January 1 in Table 2. So, children who are born on January 1st and who comply with the rules will have a school starting age that is one year higher (and one extra

Table 1: *Descriptive Statistics*

	Mean	SD	N
— Age 7 Wave —			
School starting age (years)	6.31	0.59	54,251
School starting age $\geq$ 6.5	0.22	0.41	54,251
Born in spring	0.50	0.50	54,251
Distance (in days) to January 1	-2.69	108.49	54,251
Years of schooling, highest among parents	15.51	1.99	54,148
Parents gross income	737.61	463.83	54,148
Mother's age when child was born	30.64	4.21	54,140
Birthweight (gr.)	3560.36	592.11	53,726
Female	0.49	0.50	53,726
5min APGAR score	9.79	1.06	54,231
Age	7.16	0.13	54,251
Total Difficulties	-0.01	0.99	54,251
Emotional Symptoms	-0.01	0.99	54,251
Conduct Problems	-0.01	1.00	54,251
Inattention/Hyperactivity	-0.01	0.99	54,251
Peer Problems	-0.01	0.98	54,251
Pro-social Behavior	0.01	0.99	54,251
— Age 11 Wave —			
School starting age,(years)	6.27	0.58	35,902
School starting age $\geq$ 6.5	0.19	0.39	35,902
Born in spring	0.50	0.50	35,902
Distance (in days) to January 1	-2.66	108.97	35,902
Years of schooling, highest among parents	15.68	1.95	35,825
Parents gross income	747.03	478.18	35,825
Mother's age when child was born	30.83	4.16	35,821
Birthweight (gr.)	3574.39	582.07	35,555
Female	0.50	0.50	35,555
5min APGAR score	9.79	1.06	35,888
Age	11.35	0.56	35,902
Total Difficulties	-0.03	0.98	35,902
Emotional Symptoms	-0.02	0.99	35,902
Conduct Problems	-0.02	0.98	35,902
Inattention/Hyperactivity	-0.02	0.99	35,902
Peer Problems	-0.03	0.97	35,902
Pro-social Behavior	0.01	0.98	35,902

Notes: School starting age is parent reported. Born in spring is an indicator that takes the value of 1 if the child is born in months January to June. Distance to January 1 is measure din days for the year going from July to June (January 1=0). Non-western origin is an indicator that takes the value of 1 if at least one of the parents is an immigrant from a non-western country (according to Statistics Denmark classification). Parental characteristics are measured one year after birth for children born in spring and two years after birth for children born in fall. SDQ scores are standardized by wave, before selecting on non-missing school starting age.

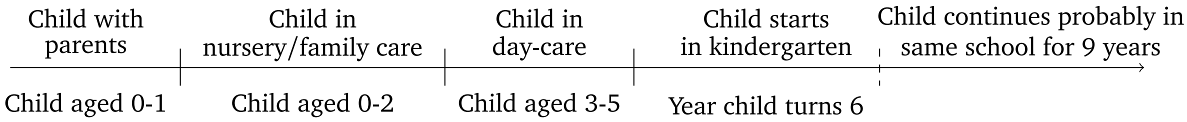


Figure 2: *Timing of childhood*

year of daycare) relative to the children born just one day earlier. However, compliance with this rule is not mandatory. That is, it is possible to postpone enrollment in school. However, this requires some effort of the parents, including meeting with representatives from the future school and the municipality administration. Contingent on individual evaluations, children may also enroll in grade zero one year *earlier* (i.e., if their birthday is before October 1). Kindergarten class is part of the primary school and free of charge in the public schools.

Table 2: *Timing of childhood for a child born December 31 and a child born January 1*

Born	December 31st	January 1st
With parents	Months 0-12	Months 0-12
In nursery	Months 13-36	Months 13-36
In day-care	Months 37-66	Months 37-78
Enroll in grade zero	Month 67	Month 79

The kindergarten class year starts with an obligatory assessment of the child's verbal communication skills and the outlining of an individual teaching plan (in Danish, Elevplan). Schools assign students to classes based on both pedagogical and practical considerations (e.g., peer composition, class-size requirements), and the principle is the same or grades one to seven. Kindergarten class has a formally specified curriculum by the Ministry of Education. The curriculum includes topics such as verbal and non-verbal communication, as well as science and nature (The Danish Ministry of Education, 2009). The Ministry of Education also specifies a minimum number of 600 teaching hours per school year (approximately 3 hours per school day). Interestingly, as almost all children attend daycare before they enroll in school and the attributes of this daycare are centrally defined, the control condition in our RD design is quite homogenous. However, the amount of time pre-school children spend in daycare varies.

4.2 Regression Discontinuity (RD) Design

Our broad question of interest involves how school starting age (SSA) influences the SDQ-based measures of mental health (Y) for individual i with covariates $\boldsymbol{X_i}$. We represent this by the following linear specification:

$$Y_i = \beta_0 + \beta_1 SSA_i + \phi \boldsymbol{X_i} + e_i \tag{1}$$

Credibly identifying the causal effect of school starting age on these outcomes is challenging because parents are likely to make decisions about when their child begins school based on information unobserved by researchers. In particular, parents who know their children face developmental challenges may be more likely to delay their child's initiation of formal schooling (i.e., negative selection into treatment). Naive OLS estimates of (1) are consistent with this concern. For example, OLS estimates suggest that children who start school late have substantially *higher* levels of inattention/hyperactivity.

We seek to identify the causal effect of SSA by leveraging the variation created by the Danish rule that children are supposed to enroll in school the year they turn six. That is, we implement an RD design that exploits the "jump" in SSA that occurs for children born January 1st or later relative to those born earlier. So, the forcing variable in this RD design (i.e., day_i) is the child's exact birth date relative to the January 1st cutoff.[14] Our reduced-form equation of interest models the SDQ-based outcomes as a flexible function of this forcing variable and a "jump" at the policy-induced threshold:

$$Y_i = \gamma_0 + \gamma_1 \mathbf{1}(day_i \geq 0) + g(day_i) + \boldsymbol{\rho} \boldsymbol{X_i} + \epsilon_i \tag{2}$$

Our parameter of interest is γ_1, which identifies the discrete change in subsequent child outcomes for those born January 1st or later, controlling for a smooth function of their day of birth and other observed traits. Later, we assess the robustness of our results to the choice of functional form for the forcing variable. For the full sample analysis, using a July to June sample, we also discuss the selection of a polynomial function of the forcing variable based on a graphical judgment and by comparing the Akaike Information Criteria (AIC) for various

[14] That is, this forcing variables takes on values of 0, 1, 2, etc. for children born on January 1st, 2nd, and 3rd respectively. For children born on December 31st, December 30th, December 29th, etc., the forcing variable takes on values of -1, -2, -3, etc.

specifications. We also present local linear regressions based the data within increasingly tight bandwidths around the threshold. We also report and discuss the corresponding IV estimates of β_1 from (1). These "treatment on the treated" estimates are equivalent to the ratio of our reduced-form estimates to the first-stage effects we describe below. In general, the causal warrant of such an RD design turns on whether the conditional change at the January 1st cutoff implies (i) variation in SSA and (ii) that this variation is "as good as randomized" (Lee and Lemieux, 2010). We now turn to evidence on both questions.

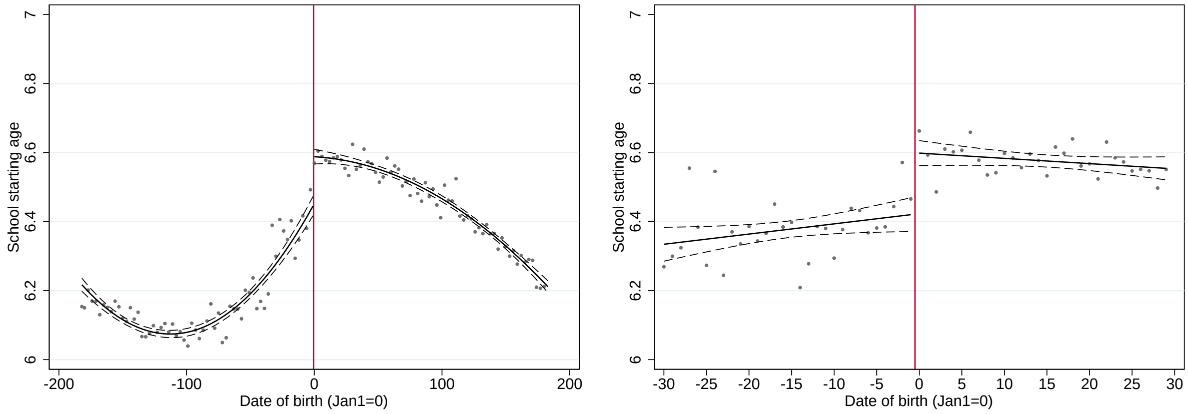
4.3 Assignment to Treatment

We first show that school starting age increases significantly for children whose birthdays are at the January 1st cutoff or later. One straightforward and unrestrictive way to show this is graphically as in Figures 3a and 3b. These figures illustrate the conditional means of school starting ages for differently sized bins defined by date of birth (i.e., 15 and 1-day bins) and for different bandwidths (i.e., the full sample and observations within 30 days of the threshold). These graphs consistently show that school starting age jumps from 6.4 to 6.6 for children born around the threshold. Interestingly, for the full-year sample in Figure 3a, the quadratic trends capture the variation in school starting age reasonably well. In the local specification in Figure 3b the linear trend seems sufficient to describe the relationship. Interestingly, this pattern implies that children born January 1st or later generally comply and begin school in August of the year they turn six. However, the compliance among children born in late December is only partial. We examine some of the issues raised by this non-compliance with respect to our "intent to treat" analysis.

We also present parametric and nonparametric estimates of this first-stage relationship based on regressions of the following form:

$$SSA_i = \delta_0 + \delta_1 \mathbf{1}(day_i \geq 0) + g(day_i) + \boldsymbol{\rho X_i} + \eta_i \tag{3}$$

Specifically, we begin with specifications that control for a linear function of the forcing variable that is allowed to vary on either side of the threshold. We then add our baseline covariates as controls as well as quadratic splines of the forcing variable. Finally, we also consider local linear regressions based only on data within a 30-day bandwidth of the threshold and conditional on a linear spline of the forcing variable as in Landersø, Nielsen,



(a) *Full year bandwidth & 3 day bins.*

(b) *30 days bandwidth & one day bins.*

Figure 3: *Date of birth and school starting age*

Table 3: *RD Estimates, First-Stage Regressions*

	(1)	(2)	(3)	(4)	(5)
Age 7 wave	0.38***	0.39***	0.17***	0.18***	0.20***
	(0.01)	(0.01)	(0.02)	(0.03)	(0.03)
Age 11 wave	0.38***	0.38***	0.15***	0.18***	0.19***
	(0.01)	(0.01)	(0.02)	(0.04)	(0.04)
Sample	Full	Full	Full	Local	Local
Specification	Linear	Linear	Quadratic	Linear	Linear
Controls	No	Yes	Yes	No	Yes

Robust standard errors in parenthesis. $^{***}p < 0.01$ $^{**}p < 0.05$, $^{*}p < 0.1$. Each cell shows the estimate from a single regression. Controls included are: indicators for birth year, age at interview, parents' years of schooling, parents' gross income, mother's age at childbirth, birth weight, gender, 5 minute APGAR score, and origin. Missing values in covariates are replaced with zeros and indicators for missing variables are included.

and Simonsen (2013).

We present the results from these first-stage models in Table 3. All the point estimates across these specifications indicate that school starting age jumps by a large and statistically significant amount. However, models that use the full sample of data and condition only on linear terms for the forcing variable suggest that this effect is substantially larger. We view this as an artifact of the non-linearities evidenced in Figure 3. When we allow the forcing variable to have a non-linear relationship with school starting age, we find that SSA jumps by 0.15 to 0.17 years at the birthdate cutoff. Interestingly, local linear regressions (i.e., columns 4 and 5 in Table 3) suggest roughly similar first stage effects (i.e., 0.18 to 0.20).

4.4 Validity of the RD Design

The prior evidence demonstrates that there is a large, statistically significant jump in school starting age for children born January 1st and later. However, there are a number of reasons to be concerned that this relationship may not constitute a valid quasi-experiment. For example, a fundamental concern in any RD design is that the value of the forcing variable relative to the threshold may be systematically manipulated by those with a differential propensity for the relevant outcomes. In this setting, we might wonder whether expectant mothers either advance or delay the timing of their birth around the January 1st threshold and that the personal and family traits influencing this choice also influence child outcomes. We present two types of evidence that are consistent with the maintained hypothesis that there is no empirically meaningful manipulation of birth dates among our respondents.

First, we evaluate the distribution of births over the cutoff. Figure 4 shows the distribution of date of birth in our sample based on the test introduced by McCrary (2008). This figure indicates that the number of births are smoothly distributed around the threshold. The null hypothesis of no jump at this threshold cannot be rejected. Interestingly, there appears to be a small drop in births around the new year (i.e., both December 31st and January 1st), which may reflect some effort to avoid giving birth during a holiday. To consider possible issues related to undiagnosed "heaping" of the forcing variable, we also show in Figure 5 a histogram of birth dates local to the threshold. These data also suggest that the frequency of observations is continuous through the threshold that defines our intent to treat.

Second, we use auxiliary regressions (i.e., the same specification as our RD design but with baseline covariates as the dependent variables) to examine the balance of observed traits of children and their families around the threshold. If the variation in school starting ages around this threshold is "as good as randomized," we would expect the pre-determined and observed traits of survey respondents to be similar on both sides of the threshold (i.e., no "jump" indicated by the RD estimates). In the appendix, Table A.2 shows these results for each of the covariates. None of the the covariates show signs of jumps at the cutoffs in either the 30 day local specification or in the full sample parametric specification. An alternative strategy for testing covariate balance is to first regress the outcome variable on all covariates and compute the predicted values. These predicted value represents an index of all the covariates that are weighted by their OLS-estimated outcome relevance. In Table 4

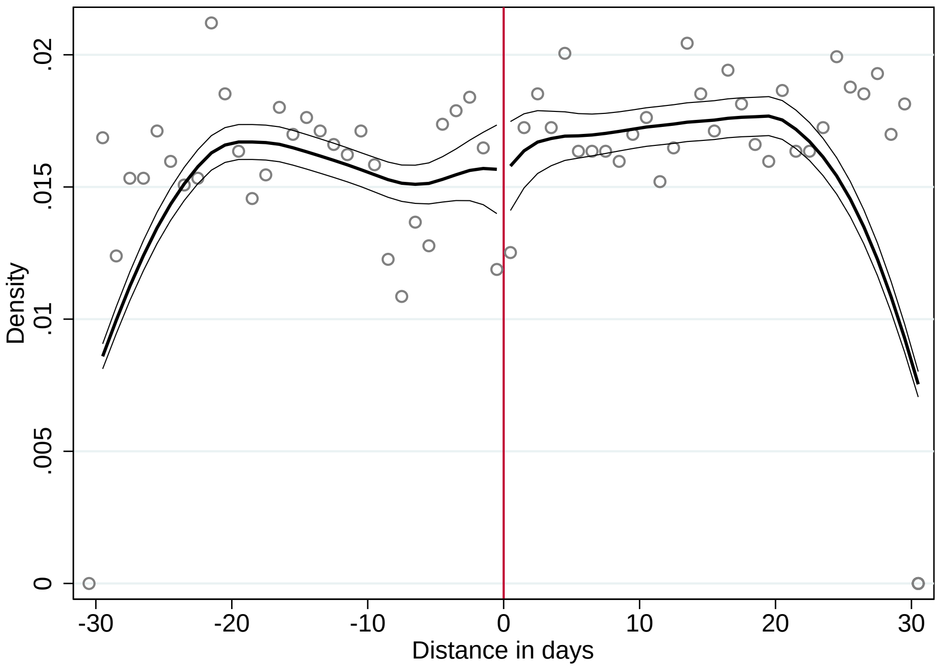


Figure 4: *Observations by date of birth, based on McCrary (2008). The jump is estimated to be -0.013 (0.086) using a one day bin width and the default bandwidth.*

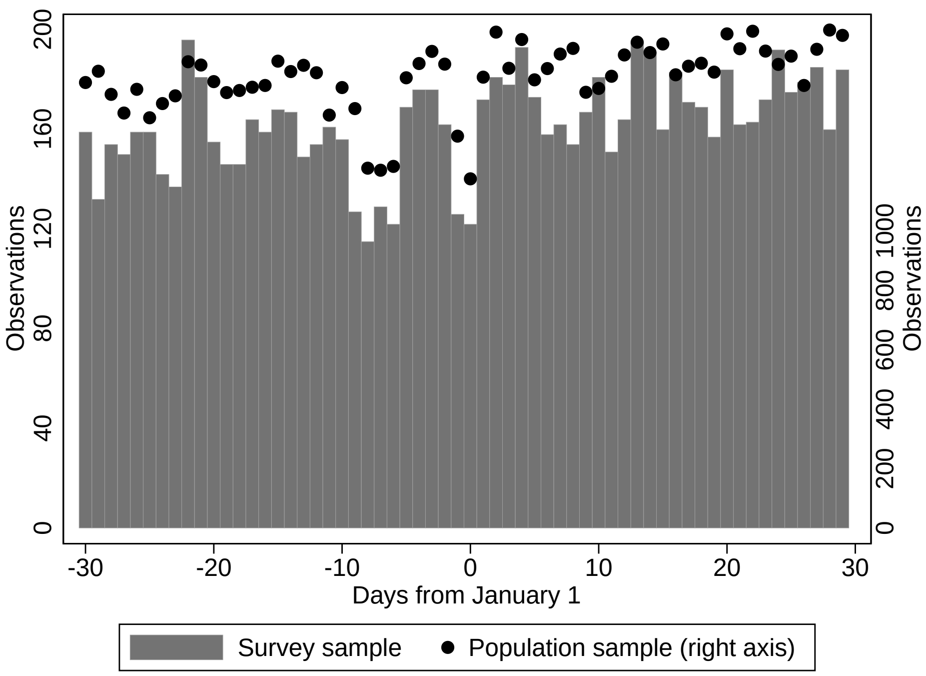


Figure 5: *Observations by date of birth, survey data and population data. The survey data is the data used in our analysis, and the population includes all children born in Denmark in the period 1998-2003.*

we show the outcome of regressing this weighted average on the cutoff and time trends for for each of the six dependent variables. As with the single-covariate regressions, there is no sign of a jump in any of these specifications.[15] The balance of outcome-relevant covariates around the January 1st threshold not only suggests a lack of manipulation of birth dates but it is also general evidence for the validity of the RD design. We should also note that we also compared the balance of several developmental variables defined for the DNBC respondents *before* they attended kindergarten (e.g., making word sounds at 18 months). We found that these traits were balanced around the threshold (Table A.3).

Table 4: *Auxiliary RD estimates, balancing of the covariates.*

	Age 7		Age 11	
	(1)	(2)	(3)	(4)
$\hat{Y}$ (Total Difficulties)	-0.008	-0.013	-0.009	-0.007
	(0.007)	(0.012)	(0.006)	(0.011)
$\hat{Y}$ (Emotional Symptoms)	-0.006	-0.012	-0.007	-0.006
	(0.005)	(0.009)	(0.004)	(0.008)
$\hat{Y}$ (Conduct Problems)	-0.005	-0.012	-0.007	-0.006
	(0.005)	(0.008)	(0.004)	(0.008)
$\hat{Y}$ (Hyperactivity)	-0.006	-0.008	-0.011	-0.005
	(0.007)	(0.011)	(0.007)	(0.012)
$\hat{Y}$ (Peer Problems)	-0.005	-0.005	-0.000	-0.004
	(0.004)	(0.008)	(0.003)	(0.006)
$\hat{Y}$ (Pro-social Behavior)	0.001	-0.008	0.002	-0.013
	(0.005)	(0.009)	(0.005)	(0.010)
Sample	Full	Local	Full	Local
Specification	Quadratic	Linear	Quadratic	Linear

Robust standard errors in parenthesis. $^{***}p < 0.01$ $^{**}p < 0.05$, $^{*}p < 0.1$. Each cell shows the estimate from a single regression. We first regress the outcome variables (in parenthesis) of the following set of covariates: indicators for birth year, age at interview, parents' years of schooling, parents' gross income, mother's age at childbirth, birth weight, gender, 5 minute APGAR score, and origin. We regress the predicted variable on an indicator for being born on January 1 or later, as well as the splines indicated by the bottom row.

Another fundamental concern with any RD design involves whether the functional form is correctly specified. In particular, a failure to specify the functional form correctly could lead to biased inferences about the true effects of our intent to treat. A visual inspection of our results provides one important and unrestrictive way to assess this concern. However, to examine the empirical relevance of functional-form issues more directly, we also report

[15]Note that both Table A.2 and Table 4 show uncorrected standard errors and significance levels. Any corrections for multiple testing will make the conclusions of no correlation even stronger.

results from specifications that add quadratic terms for the forcing variable. We also consider the corresponding information criteria across specifications. And we report estimates based on samples within increasingly tight bandwidths around the threshold (i.e., local linear regressions). Whether our results are robust to these specification choices speaks to the relevance of functional-form issues.

At least two other internal-validity concerns are unique to our application and merit scrutiny. First, as noted early, our treatment contrast necessarily conflates higher school starting ages with fewer years of schooling at the time of observation. That is, our intent-to-treat (i.e., a birth date of January 1st or later) implies *both* a higher school starting age and fewer years of formal schooling at the time parents rate their children on the SDQ. However, there are several reasons to deprecate the role of years of schooling in our analysis. For example, our pattern of results (i.e., effects on only one SDQ construct and not on the other measures of psychological adaptation) are not easily reconcilable with effects due to years of schooling but are consistent with the theorized effects of higher school starting ages. Furthermore, we find that our results are quite similar in size and significance among children at age 7 as at age 11 when the differences in years of schooling are relatively smaller. This pattern would only be consistent with effects due to years of schooling if a year has an additive effect without fade-out. Also, given that years of schooling are likely to have a positive effect on our mental-health measures (at least in later childhood), the collinearity in these measures (higher school starting age and *fewer* years of schooling) would not imply a bias that is problematic for our main findings.[16].

A second internal-validity threat unique to our setting involves reference biases in the SDQ ratings. It may be that children whose schooling is delayed are more likely to be rated positively simply because they appear to have better psychological adaptations than their younger classroom peers. Indeed, there is provocative evidence among U.S. children (Elder, 2010) that teachers are significantly more likely to rate children who are young for their grade as having ADHD. However, Elder (2010) finds that *parental* assessments (i.e., like those in the DNBC) are not subject to these biases; in all likelihood, because they have different reference points than teachers. Moreover, if the parent reports in the DNBC were subject to such biases, we would also expect to find effects on SDQ constructs other than

[16]A study by Leuven, Lindahl, Oosterbeek, and Webbink (2010) utilizes the unique rolling-admissions policies in the Netherlands and their interaction with school holidays, and finds that earlier enrollment opportunities improve the test performance of disadvantaged students but have no or possibly negative effects of more advantaged students

inattention/hyperactivity but do not. We also hypothesized that, if our results were sensitive to rater biases, they would be attenuated among children who have older siblings (i.e., a different reference point for parents). However, we find that they are not.

In sum, we find broad support for the internal validity of our research design. However, our analysis, like most RD applications, is qualified by several caveats related to external validity. First, because our estimates are defined by variation around the January 1st threshold, they are necessarily local estimates. Whether our results generalize to those born at other times is uncertain. There is evidence shows that season of birth is not random with respect to parental characteristics (Buckles and Hungerman, 2013) so the localness of our RD estimates may have some empirical salience. Second, our estimates are qualified by the non-random non-response to the last DNBC survey waves. In general, these respondents tended to be more affluent. A third concern is related to the "fuzzy" nature of our RD design. If our treatment effects of interest are not homogenous, the LATE theorem implies that our treatment estimates are defined for the sub-population of "compliers" with their intent to treat (Imbens and Angrist, 1994). We speak to these concerns in two ways. One is to estimate our treatment effects separately for sub-samples of the data defined by pre-treatment characteristics (e.g., boys versus girls). Second, using a straightfoward technique recently introduced by Bertanha and Imbens (2014) we examine whether our complier population is distinctive.

5 Results

5.1 Graphical Evidence

We begin with an unrestrictive, visual representation of our reduced-form results. First, Figure 6 shows, for each distinct SDQ measure observed at age 7, the conditional means by day of birth (i.e., in 3-day bins) on each side of the January 1st threshold. The first panel of this figure shows a distinct drop in the total difficulties score (i.e., of roughly 0.15 SD) for children whose birthday is January 1st and later. The next four panels (i.e., b through e) suggest that this drop occurred for each of the four measures that constitute the difficulties score. However, the decrease in difficulties is uniquely large for the inattention/hyperactivity measure (i.e., the measure indicating a lack of self regulation). Panel (f) suggests that there is a noticeable increase in the pro-social measure for children born January 1st or later.

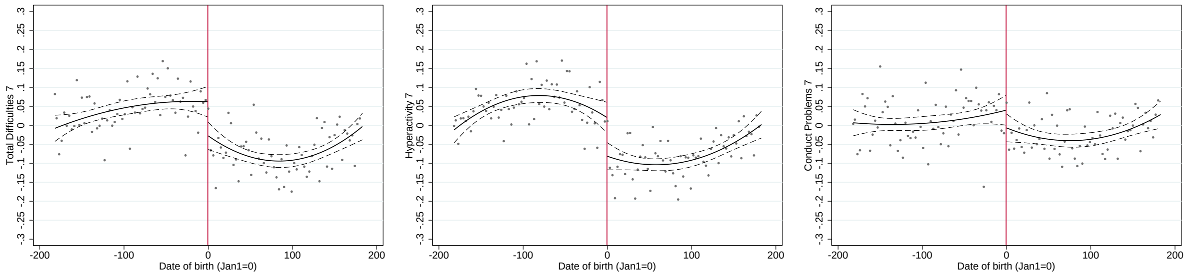
These age-7 results provide clear evidence that quasi-random assignment to a delayed school start appears to improve mental health, particularly self-regulation, reported at age 7. However, one concern with these short-run findings is that they may be an artifact of the age at which parents report these data. In particular, the children for whom the intent to treat (ITT) is one (i.e., those born January 1st or later) are more likely to be in kindergarten (i.e., a half-day program) relative to the ITT=0 children who are more likely to be in 1st grade, a full-day program. So, it is possible that these effects, while valid, reflect the current differences in the student's exposure to formal schooling rather than deeper developmental effects. The fact that the effects are concentrated in self-regulation rather than other constructs (as well as the evidence of *positive* effect on sociabilty) argues somewhat against this interpretation.

However, a more compelling way to address this concern is to consider outcomes at a later age when the children, regardless of their ITT, have long spells of formal schooling. In Figure 7, we show such evidence by illustrating the mean values of the SDQ measures across 3-day bins defined by date of birth for children observed in the most recent age-11 wave of the DNBC. As with the age-7 data, these graphs suggest that those born on or after the cutoff (i.e., those with an ITT to delay their school start) have substantially lower levels of difficulties and a higher level of sociability. Again, we see (i.e., panel (b) in Figure 7 that this effect is uniquely large with respect to the inattention/hyperactivity construct.

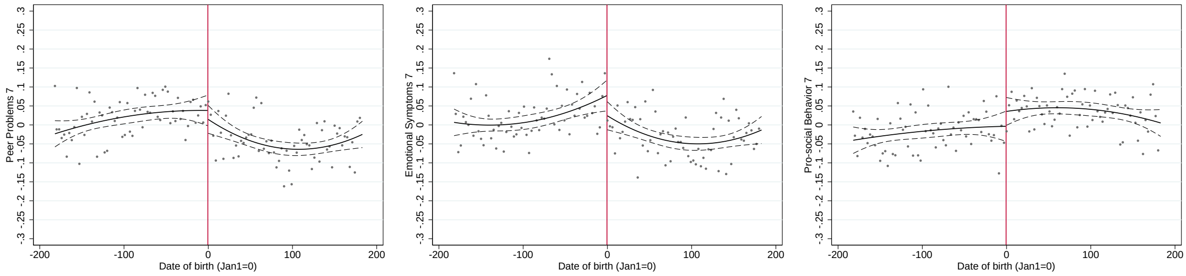
5.2 Main Estimates

Our graphical results provide highly suggestive evidence that a higher school starting age leads to an improvement in children's mental health, particularly with respect to self regulation. In this section, we present our key RD estimates. This regression framework allows to identify the point estimates of interest and, critically, test their statistical significance. However, this framework also allows us to explore the robustness of our visual evidence.

In Table 5, we present the reduced-form RD estimates for age-7 SDQ measures across five different specifications. The first two specifications condition on a linear spline of the forcing variable and one includes controls for our baseline covariates. These results suggest that the ITT generates statistically significant reductions in each difficulty and a statistically significant increase in the pro-social construct. However, our graphical evidence suggests that the relationship between date of birth and the SDQ measures is not always linear. In

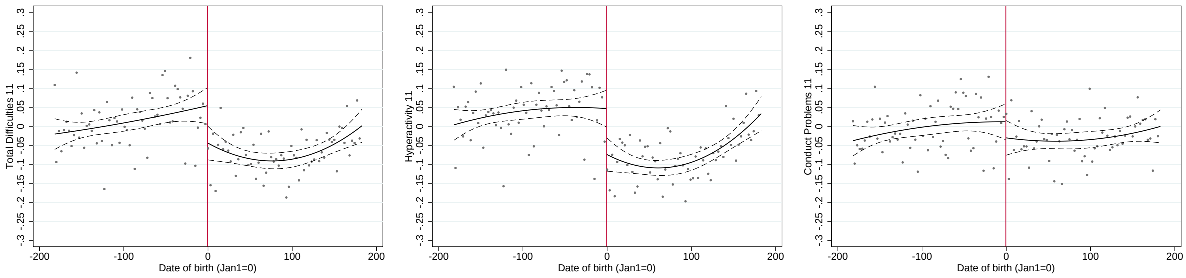


(a) *Total Difficulties* **(b)** *Inattention/Hyperactivity* **(c)** *Conduct*

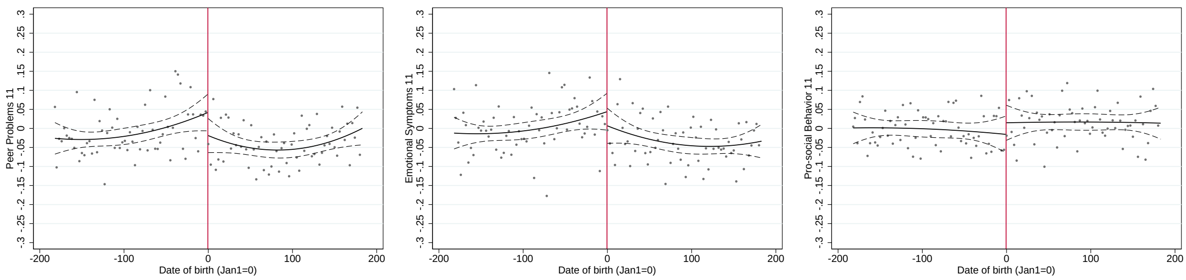


(d) *Peer problems* **(e)** *Emotional* **(f)** *Pro-social*

Figure 6: *Reduced form relationship, age 7. Bin width: 3 days. Quadratic fits.*



(a) *Total Difficulties* **(b)** *Inattention/Hyperactivity* **(c)** *Conduct*



(d) *Peer problems* **(e)** *Emotional* **(f)** *Pro-social*

Figure 7: *Reduced form relationship, age 11. Bin width: 3 days. Quadratic fits.*

our third specification, we condition on both linear and quadratic versions of the forcing variable, while still allowing these variables to vary on both sides of the cutoff. In this

Table 5: *Reduced-form RD Estimates, The Effect of $day_i \geq 0$ on SDQ at age 7*

	(1)	(2)	(3)	(4)	(5)
Total Difficulties	-0.16***	-0.15***	-0.08***	-0.14***	-0.12***
	(0.02)	(0.02)	(0.03)	(0.05)	(0.05)
Emotional Symptoms	-0.07***	-0.06***	-0.05*	-0.08*	-0.07
	(0.02)	(0.02)	(0.03)	(0.05)	(0.05)
Conduct Problems	-0.07***	-0.06***	-0.04	-0.06	-0.05
	(0.02)	(0.02)	(0.03)	(0.05)	(0.05)
Inattention/Hyperactivity	-0.19***	-0.19***	-0.09***	-0.15***	-0.14***
	(0.02)	(0.02)	(0.03)	(0.05)	(0.05)
Peer Problems	-0.08***	-0.07***	-0.01	-0.04	-0.04
	(0.02)	(0.02)	(0.03)	(0.05)	(0.05)
Pro-social Behavior	0.05***	0.05***	0.04	0.09*	0.09**
	(0.02)	(0.02)	(0.03)	(0.05)	(0.05)
Observations	54,251	54,251	54,251	7,642	7,642
Sample	Full	Full	Full	Local	Local
Specification	Linear	Linear	Quadratic	Linear	Linear
Controls	No	No	Yes	No	Yes

Robust standard errors in parenthesis. $^{***}p < 0.01$ $^{**}p < 0.05$, $^{*}p < 0.1$. Each cell shows the estimate from a single regression. Controls included are: indicators for birth year, age at interview, parents' years of schooling, parents' gross income, mother's age at childbirth, birth weight, gender, 5 minute APGAR score, and origin. Missing values in covariates are replaced with zeros and indicators for missing variables are included.

specification, our RD estimates are generally smaller. And we find that the only statistically significant reduction implied by the ITT is in the inattention/hyperactivity construct (effect size = -0.09). The reduction in emotional problems (effect size = -0.05) is only weakly significant. When using the full sample, we prefer to condition on both linear and quadratic terms for the forcing variable because of the graphical evidence but also because Akaike's Information Criteria (AIC) generally privilege this specification. In the final two columns of Table 5, we report non-parametric estimates based on an unweighted local linear regression (LLR) using observations in a one-month bandwidth around the cutoff (i.e., December and January births). These LLR results similarly indicate that the effect of the January 1st cutoff is effectively concentrated in the inattention/hyperactivity measure.[17]

In Table 6, we present similarly constructed reduced-form RD estimates for age-11 SDQ measures. As with the age-7 results, the full-sample results that condition on linear terms for the forcing variable suggest that the ITT led to large, sustained reductions in difficulties.

[17] We note that we have not formally applied multiple-comparison adjustments to our inferences. However, our main results are estimated with sufficient precision that they would remain statistically significant after correcting for examining 12 core outcomes (i.e., 6 SDQ measures across two age groups).

However, more flexible functional forms (i.e., quadratic terms for the forcing variable and the LLR approach) again indicate that the effects unique to birth-date cutoff are concentrated in the inattention/hyperactivity score. Overall, these point estimates indicate that a delayed school starting age causes a significant improvement in self-regulation that is sustained for at least several years and also qualitatively large. It should be noted that these ITT estimates identify the change in self-regulation implied by the change in school starting age from our first-stage equations (i.e., roughly 0.2 years).

Our implied estimate of the effect of a full year increase in school starting age is 5 times as large as these reduced-form effects. For example, using the LLR results conditional on controls, we find that increasing the school starting age by one year reduces inattention/hyperactivity at age 7 by 0.7 SD (i.e., -0.14/0.20). The corresponding 2SLS estimate for age 11 is -0.68 SD (i.e., -0.13/0.19). Arguably, these effect sizes are quite large, particularly for at-scale field settings.

Another potentially useful way to benchmark effects this large is to benchmark them against the mental-health gaps observed in the data. For example, children from families in the lowest decile of income have inattention/hyperactivity scores that are 0.61 SD higher at age 7 and 0.5 SD higher at age 11 relative to children in the top decile. And boys have higher inattention/hyperactivity scores than girls (i.e., generally about 0.7 SD). Our finding indicates that a one-year increase in school starting age produces an effect that is as large or larger than these mental-health gaps by income and gender.

One heuristic indication of the robustness of our findings is the correspondence between our visual evidence, our full-sample specifications that include quadratic terms and our non-parametric estimates based on local linear regressions. To explore this robustness further, we constructed IV/2SLS estimates based on our RD design for multiple bandwidths that use only the data within increasingly tight bandwidths around the cutoff (i.e., the full sample to bandwidths as tight as 20 days). We present these results visually in Figure 8 and Figure 9, respectively for the age-7 and age-11 samples. In this figures, the three horizontal lines indicate the full-sample point estimates based on specifications that add higher-order polynomials of the forcing variable (i.e., linear, quadratic, and cubic). The dark dots and error bars provide the point estimates and 95-percent confidence intervals from LLR specifications based on increasingly tight bandwidths. Unsurprisingly, all our models lose statistical precision when the sample sizes are reduced. However, our main inattention/hyperactivity

Table 6: *Reduced-form RD Estimates, The Effect of $day_i \geq 0$ on SDQ at age 11*

	(1)	(2)	(3)	(4)	(5)
Total Difficulties	-0.14***	-0.14***	-0.09***	-0.10*	-0.09*
	(0.02)	(0.02)	(0.03)	(0.05)	(0.05)
Emotional Symptoms	-0.04**	-0.04*	-0.03	-0.02	-0.01
	(0.02)	(0.02)	(0.03)	(0.06)	(0.06)
Conduct Problems	-0.07***	-0.06***	-0.03	-0.00	0.01
	(0.02)	(0.02)	(0.03)	(0.06)	(0.06)
Inattention/Hyperactivity	-0.19***	-0.18***	-0.11***	-0.14***	-0.13**
	(0.02)	(0.02)	(0.03)	(0.06)	(0.05)
Peer Problems	-0.07***	-0.08***	-0.06*	-0.09*	-0.09
	(0.02)	(0.02)	(0.03)	(0.06)	(0.05)
Pro-social Behavior	0.03	0.02	0.03	0.02	0.03
	(0.02)	(0.02)	(0.03)	(0.06)	(0.06)
Observations	35,902	35,902	35,902	5,050	5,050
Sample	Full	Full	Full	Local	Local
Specification	Linear	Linear	Quadratic	Linear	Linear
Controls	No	No	Yes	No	Yes

Robust standard errors in parenthesis. $^{***}p < 0.01$ $^{**}p < 0.05$, $^{*}p < 0.1$. Each cell shows the estimate from a single regression. Controls included are: indicators for birth year, age at interview, parents' years of schooling, parents' gross income, mother's age at childbirth, birth weight, gender, 5 minute APGAR score, and origin. Missing values in covariates are replaced with zeros and indicators for missing variables are included.

result is quite robust. The point estimates remain large in absolute value and actually, tend to get larger using only observations close to the threshold. And, across nearly all of these specifications, there is sufficient precision to reject a null hypothesis of no effect.

Table 7: *2SLS estimates, the effect of School Starting Age on SDQ at age 7*

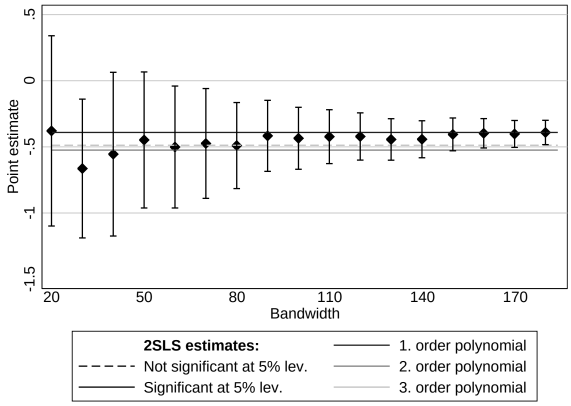
	(1) Emotional	(2) Conduct	(3) Inattention/ Hyperactivity	(4) Peer Problems	(5) Pro-social
Main	-0.27	-0.24	-0.53***	-0.09	0.26
	(0.17)	(0.16)	(0.17)	(0.17)	(0.16)
Boys	-0.12	-0.42	-0.83	0.35	0.39
	(0.57)	(0.61)	(0.66)	(0.62)	(0.62)
Girls	-0.32**	-0.19	-0.45***	-0.21	0.23*
	(0.15)	(0.14)	(0.14)	(0.13)	(0.13)
Highly educated	-0.46*	-0.40	-0.84***	-0.39	0.14
	(0.26)	(0.25)	(0.28)	(0.25)	(0.25)
Low educated	-0.13	-0.12	-0.30	0.14	0.36*
	(0.22)	(0.22)	(0.22)	(0.23)	(0.21)
High income	-0.51**	-0.45*	-0.70***	-0.34	0.22
	(0.23)	(0.23)	(0.24)	(0.22)	(0.23)
Low income	-0.04	-0.06	-0.38	0.15	0.31
	(0.25)	(0.24)	(0.24)	(0.25)	(0.23)
Low birthweight	-0.31	-0.05	-0.51**	-0.07	0.26
	(0.24)	(0.24)	(0.24)	(0.24)	(0.23)
High birthweight	-0.20	-0.40*	-0.53**	-0.09	0.27
	(0.22)	(0.22)	(0.23)	(0.22)	(0.22)
No older sibs	-0.06	-0.07	-0.35*	0.01	0.03
	(0.22)	(0.20)	(0.21)	(0.21)	(0.20)
Older sibs	-0.52*	-0.51*	-0.81***	-0.28	0.60**
	(0.27)	(0.29)	(0.30)	(0.28)	(0.29)

All results are based on the full sample using quadratic splines. Robust standard errors in parenthesis. $^{***}p < 0.01$ $^{**}p < 0.05$, $^{*}p < 0.1$. Each cell shows the estimate from a single regression. Controls included are: indicators for birth year, age at interview, parents' years of schooling, parents' gross income, mother's age at childbirth, birth weight, gender, 5 minute APGAR score, and origin. Missing values in covariates are replaced with zeros and indicators for missing variables are included. All sample splits are done at the median. Non-singletons are always defined as having an older sibling.

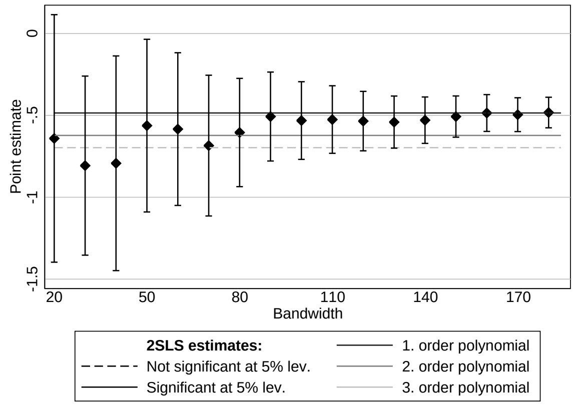
Table 8: *2SLS estimates, the effect of School Starting Age on SDQ at age 11*

	(1)	(2)	(3)	(4)	(5)
	Emotional	Conduct	Inattention/ Hyperactivity	Peer Problems	Pro-social
Main	-0.18	-0.22	-0.72***	-0.40*	0.19
	(0.22)	(0.22)	(0.23)	(0.22)	(0.22)
Boys	1.21	-0.53	-1.57	-0.99	-0.10
	(1.39)	(1.26)	(1.67)	(1.42)	(1.25)
Girls	-0.39**	-0.18	-0.59***	-0.30*	0.23
	(0.19)	(0.17)	(0.17)	(0.17)	(0.16)
Highly educated	-0.15	-0.36	-1.19**	-0.44	0.31
	(0.38)	(0.38)	(0.49)	(0.39)	(0.41)
Low educated	-0.17	-0.11	-0.41	-0.36	0.11
	(0.26)	(0.26)	(0.26)	(0.26)	(0.25)
High income	-0.44	-0.52	-1.15***	-0.63*	0.23
	(0.32)	(0.32)	(0.39)	(0.34)	(0.33)
Low income	0.09	0.07	-0.35	-0.16	0.12
	(0.30)	(0.31)	(0.30)	(0.30)	(0.29)
Low birthweight	0.20	0.14	-0.49	0.04	0.22
	(0.32)	(0.31)	(0.32)	(0.31)	(0.31)
High birthweight	-0.54*	-0.56*	-0.93***	-0.81**	0.16
	(0.30)	(0.31)	(0.33)	(0.33)	(0.30)
No older sibs	0.17	0.04	-0.27	-0.14	-0.28
	(0.28)	(0.27)	(0.27)	(0.28)	(0.28)
Older sibs	-0.63*	-0.57	-1.35***	-0.72*	0.75*
	(0.37)	(0.37)	(0.47)	(0.39)	(0.39)

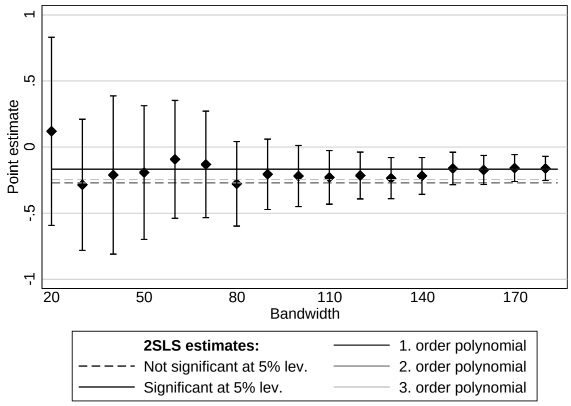
All results are based on the full sample using quadratic splines. Robust standard errors in parenthesis. $^{***}p < 0.01$ $^{**}p < 0.05$, $^{*}p < 0.1$. Each cell shows the estimate from a single regression. Controls included are: indicators for birth year, age at interview, parents' years of schooling, parents' gross income, mother's age at childbirth, birth weight, gender, 5 minute APGAR score, and origin. Missing values in covariates are replaced with zeros and indicators for missing variables are included. All sample splits are done at the median. Non-singletons are always defined as having an older sibling.



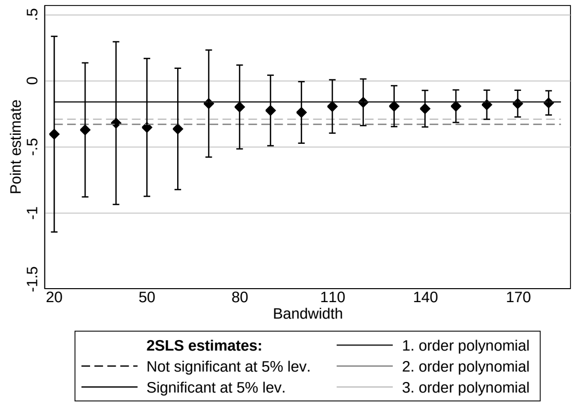
(a) *Total difficulties*



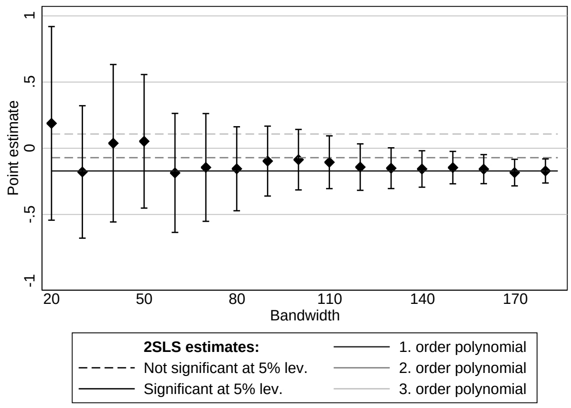
(b) *Inattention/Hyperactivity*



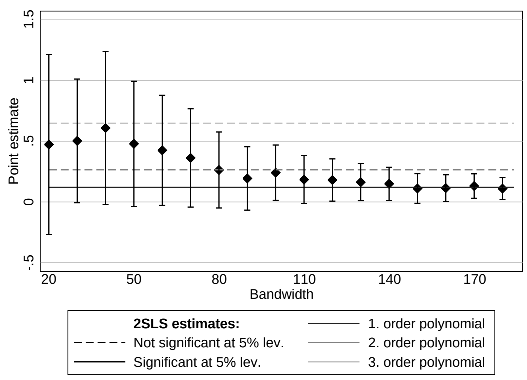
(c) *Conduct*



(d) *Emotional symptoms*

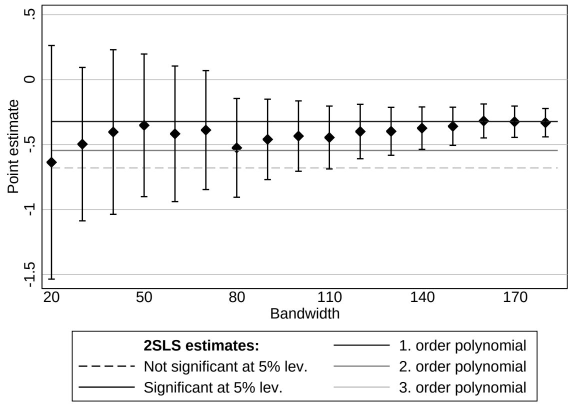


(e) *Peer problems*

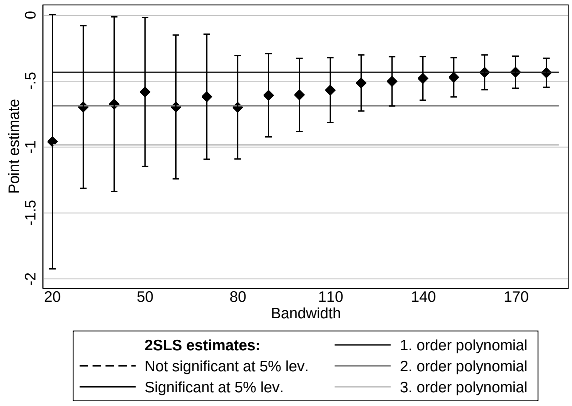


(f) *Pro-social*

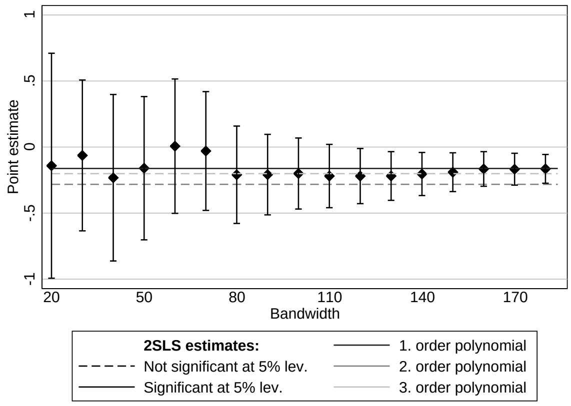
Figure 8: *Bandwidth sensitivity, age 7. Each diamond marker is the 2SLS point estimate from a local regression with the bandwidth size denoted on the x-axis. The bandwidth size increases in steps of 10 days. A bandwidth of 10 implies a sample of children born 10 days before and after January 1st. The horizontal lines are the 2SLS point estimate from a regression using the full sample with separate trends on each side of the January 1st cutoff. The lines are solid if the estimate is significant on a five percent level, and dashed if it is not significant on a five percent level.*



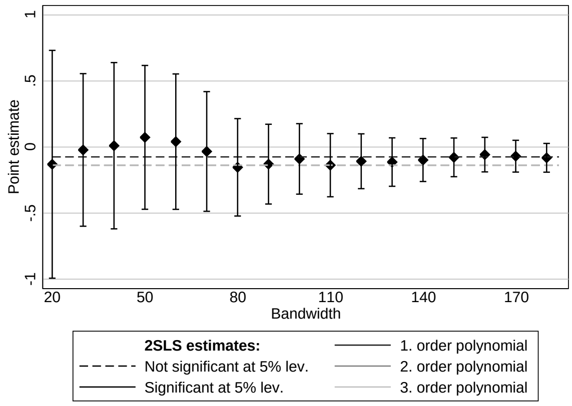
(a) *Total difficulties*



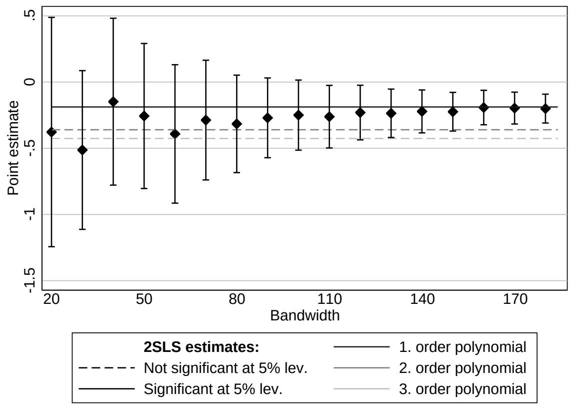
(b) *Inattention/Hyperactivity*



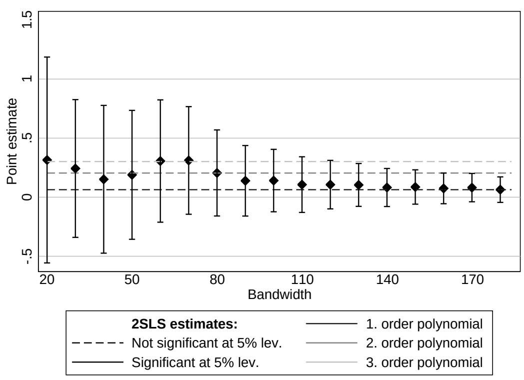
(c) *Conduct*



(d) *Emotional symptoms*



(e) *Peer problems*



(f) *Pro-social*

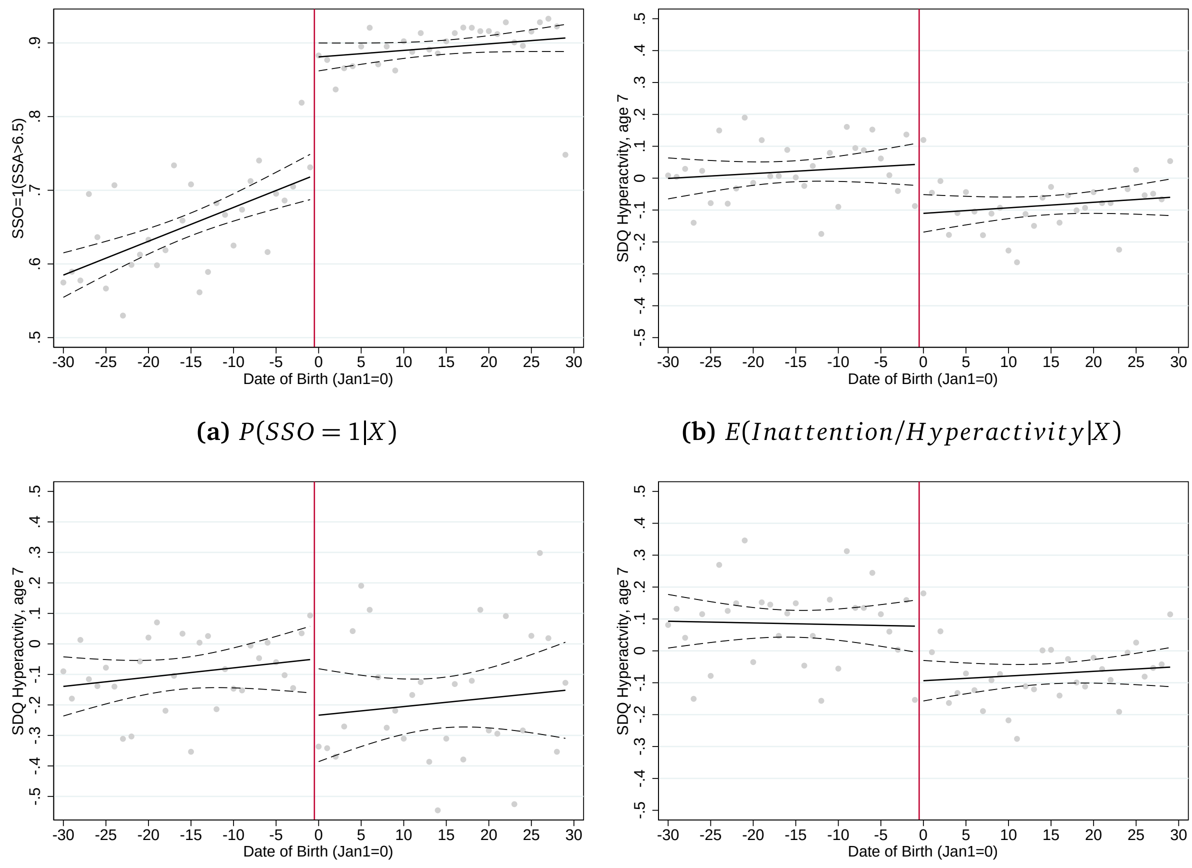
Figure 9: *Bandwidth sensitivity, age 11. Each diamond marker is the 2SLS point estimate from a local regression with the bandwidth size denoted on the x-axis. The bandwidth size increases in steps of 10 days. A bandwidth of 10 implies a sample of children born 10 days before and after January 1st. The horizontal lines are the 2SLS point estimate from a regression using the full sample with separate trends on each side of the January 1st cutoff. The lines are solid if the estimate is significant on a five percent level, and dashed if it is not significant on a five percent level.*

5.3 Treatment Heterogeneity

Our main RD results provide robust evidence that a higher school starting age leads to a large and persistent increase in one particular dimension of children's mental health (i.e., self-regulation). However, there are several ways in which the generalizability of this evidence may be limited. For example, both local nature of an RD estimand and the non-random participation of DNBC respondents to the last two survey waves raise external-validity concerns. Additionally, because we have a "fuzzy" RD design, the LATE theorem (Imbens and Angrist, 1994) implies that, in the absence of constant treatment effects, our point estimates are defined for the subpopulation of "compliers" (i.e., those who choose a treatment condition consistent with their ITT). To examine the empirical relevance of this concern, we follow the suggestion recently introduced by Bertanha and Imbens (2014). They recommend examining the continuity of outcomes, separately for children who took up the "treatment" and those who do not.

To apply this guidance in our setting, we defined the treatment as a binary indicator for older school starting age, SSO: first entering kindergarten age 6.5 years or more. In panel (a) of Figure 10, we show for the age-7 sample that this treatment "jumps" significantly at the threshold. Panel (b) illustrates the drop in the inattention/hyperactivity measure at this threshold. Panel (c) illustrates how the self-regulation measure changes at the threshold using only observations for which $SSO = 0$. Using these data, the threshold effectively separates "compliers" and "never takers" on the left from "never takers" on the right. The discrete jump in panel (c) implies that the complier population has higher levels of inattention/hyperactivity than the never-takers (i.e., in the absence of treatment). Panel (d) presents a similarly constructed graph but using data only from those who took up the treatment (i.e., $SSO = 1$). This graph separates "always-takers" on the left from a population of always-takers and compliers on the right. The significant drop in the inattention/hyperactivity measure to the right of the threshold indicates that, even when all are taking the treatment, compliers have lower levels of inattention/hyperactivity than always-takers.

In Figure 11, we see effectively similar results when using the age-11 data. What do these results imply? We believe that they are consistent with the assertion that the complier sub-population is a distinct one that may have treatment effects that differ from those for other parts of the population. For example, it is unsurprising that those who never choose to



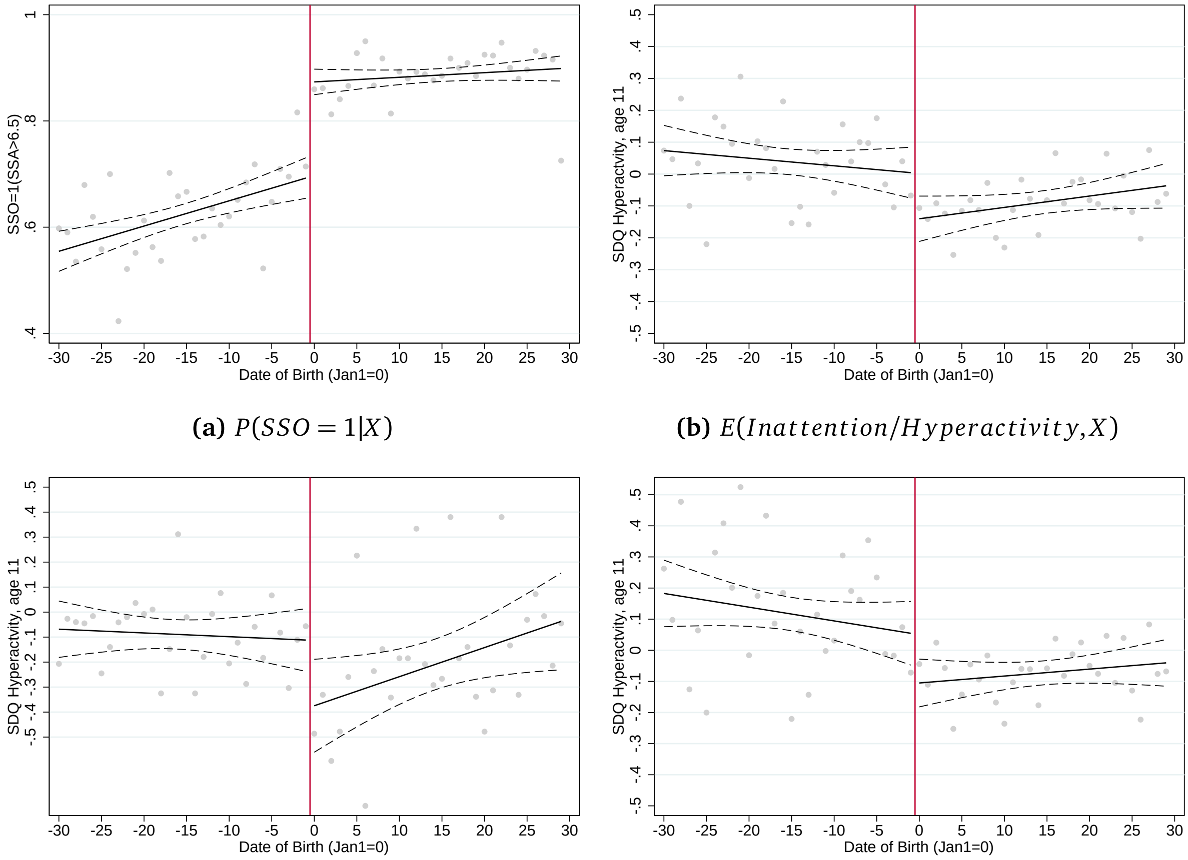
(a) $P(SSO = 1|X)$ **(b)** $E(Inattention/Hyperactivity|X)$

(c) $E(Inattention/Hyperactivity|SSO = 0, X)$ **(d)** $E(Inattention/Hyperactivity|SSO = 1, X)$

Figure 10: *Inattention/Hyperactivity at age 7, by treatment status.*

take up a delayed school start have low levels of inattention/hyperactivity (i.e., high degree of self-regulation) relative to the population that would comply when encouraged (panel c). The never-takers may rightfully see little benefit in delaying a school start. Similarly, panel (d) indicates that always-takers have uniquely high levels of inattention-hyperactivity and/or may have smaller treatment effects than compliers. This is consistent with the hypothesis that those who always seek a higher school starting age have unique developmental challenges that may be comparatively immune to the effects of a late start (i.e., relative to compliers).

To explore these issues in a more conventional and direct manner, we also examined how our key findings varied for subpopulations of the DNBC samples defined by baseline traits. Specifically, we estimated the effect of school starting age on each SDQ measure using our RD design, first, for boys and girls separately and then for respondents who were above the sample median values for education, income, and birt hweight. We report these 2SLS results in Tables 7 and 8 for the age-7 and age-11 samples, respectively.



(a) $P(SSO = 1|X)$ **(b)** $E(Inattention/Hyperactivity, X)$

(c) $E(Inattention/Hyperactivity|SSO = 0, X)$ **(d)** $E(Inattention/Hyperactivity|SSO = 1, X)$

Figure 11: *Inattention/Hyperactivity at age 11, by treatment status.*

Interestingly, these estimates indicate that a school starting age had statistically insignificant effects for boys across all measures and both ages. However, these null findings reflect a considerable loss in precision for boys. In fact, we find that the first-stage effect for boys is smaller (0.07 compared to 0.27 for girls). So, our identifying variation is uniquely relevant for girls. And estimates based only on girls indicates that a high school starting age improves both self regulation and emotional problems. Our remaining results indicate that the mental-health benefits of a higher school starting age are almost exclusively concentrated among socioeconomically advantaged children (i.e., higher parental education, income and birthweight). These results are consistent with the hypothesis that an earlier start to formal schooling confers comparative benefits to disadvantaged children.

Another relevant type of treatment heterogeneity concerns how the effects of a delayed school start may influence more severe levels of mental illness. Our prior estimates effectively identify the changes in mean SDQ measures, which are in diagnostically normal ranges. However, as noted earlier, each SDQ score can be classified as one of three lev-

Table 9: *2SLS estimates, the Effects of $day_i \geq 0$ on Abnormal/Borderline SDQ*

	- - - Age 7 - - -				- - - Age 11 - - -			
	(1)		(2)		(3)		(4)	
	Abnormal		Borderline/Abnormal		Abnormal		Borderline	
Total Diff.	-0.09***	[0.02]	-0.04	[0.06]	-0.03	[0.03]	-0.05	[0.07]
	(0.03)		(0.05)		(0.04)		(0.05)	
Emotional Sympt.	-0.10*	[0.07]	-0.13*	[0.14]	-0.05	[0.09]	0.06	[0.16]
	(0.05)		(0.07)		(0.07)		(0.08)	
Conduct Problems	-0.00	[0.05]	-0.05	[0.14]	0.04	[0.03]	0.05	[0.09]
	(0.04)		(0.06)		(0.04)		(0.06)	
Inattention/ Hyperactivity	-0.10**	[0.05]	-0.15***	[0.08]	-0.12**	[0.05]	-0.13**	[0.08]
	(0.04)		(0.05)		(0.05)		(0.06)	
Peer Problems	-0.01	[0.04]	0.01	[0.08]	-0.06	[0.06]	-0.03	[0.12]
	(0.04)		(0.05)		(0.05)		(0.07)	
Prosocial Scale	-0.03	[0.02]	-0.05	[0.06]	-0.01	[0.02]	-0.00	[0.05]
	(0.03)		(0.04)		(0.03)		(0.05)	
Observations	54,251		54,251		35,902		35,902	

Means of the dependent variables in square-brackets. Robust standard errors in parenthesis. $^{***}p < 0.01$ $^{**}p < 0.05$, $^{*}p < 0.1$. Each cell shows the estimate from a single regression. Controls included are: indicators for birth year, age at interview, parents' years of schooling, parents' gross income, mother's age at childbirth, birth weight, gender, 5 minute APGAR score, and origin. Missing values in covariates are replaced with zeros and indicators for missing variables are included.

els: normal, borderline, and abnormal. To explore this form of heterogeneity, we estimated 2SLS models using our RD design and binary indicators for an abnormal rating (or for a borderline/abnormal rating) as the dependent variables. We report these RD estimates for the age-7 and age-11 samples in Table 9. We also report the mean value of these dependent variables. In general, diagnostically abnormal ratings on these scales are not common. For example, across both age 7 and age 11, only 5 to 8 percent of respondents had inattention/hyperactivity ratings that qualified as abnormal or borderline. Consistent with our prior results, these RD estimates indicate that a birthdate at the cutoff or later significantly reduces the probability of an abnormal inattention/hyperactivity rating. These estimated effects are also quite large and suggest that a year delay in starting school virtually eliminates the probability of an abnormal rating for the typical child.

6 Discussion and conclusions

The decision to delay the age at which children in developed nations begin formal schooling is increasingly common. These delays may confer developmental advantages through both relative and absolute-age mechanisms. However, an active research literature has generally found that these delays do not clearly result in longer-run educational or economic advantages. In this study, we examined the effect of school starting age on distinctive and more proximate outcomes: measures of mental health in childhood. One key feature of our study is the availability of data on several psychopathological constructs from a widely used and extensively validated mental-health screening tool fielded among children in the Danish National Birth Cohort (DNBC) study. We are also able to identify the causal effect of higher school starting ages by leveraging the Danish rule that children should begin kindergarten in the calendar year in which they turn six. We match the children in the DNBC to the Danish administrative registries that include the exact day of birth and confirm that school starting age increases significantly for children born after the cutoff.

The results based on this "fuzzy" regression-discontinuity design indicate that delays in school starting age imply substantial improvements in mental health (e.g., reducing the overall "difficulties" score by at least 0.5 SD). The evidence for these effects is robust and, critically, persists in the latest wave of the DNBC when the children were aged 11. However, we also find that these mental-health gains are narrowly confined to one particular construct: the inattention/hyperactivity score (i.e., a measure indicating a lack of self regulation). Interestingly, this finding is consistent with one prominent theory of why delayed school starts are beneficial. Specifically, a literature in developmental psychology emphasizes the importance of pretend play in the development of children's emotional and intellectual self-regulation. Children who delay their school staring age may have an extended (and appropriately timed) exposure to such playful environments. Our findings are consistent with this absolute-age mechanism and suggest that there may be broader developmental gains to policies that delay the initiation of formal schooling (and that support playful early-childhood environments). However, we also note that there are several external-validity caveats to our study (e.g., the localness of our RD estimands, evidence for heterogeneous treatment effects). These concerns about generalizability underscore the need for further research that can guide effectively targeted and designed programs and policies.

References

ACHENBACH, T. M., A. BECKER, M. DÖPFNER, E. HEIERVANG, V. ROESSNER, H.-C. STEINHAUSEN, AND A. ROTHENBERGER (2008): "Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ instruments: research findings, applications, and future directions," *Journal of Child Psychology and Psychiatry*, 49(3), 251–275.

ALIPRANTIS, D. (2014): "When should children start school?," *Journal of Human Capital*, 8(4), 481–536.

AMERICAN PSYCHIATRIC ASSOCIATION (1994): "Diagnostic and statistical manual of mental disorders (DCM)," Discussion paper, American Psychiatric Association, Washington, DC.

ANGRIST, J. D., AND J.-S. PISCHKE (2008): *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

BASSOK, D., AND S. F. REARDON (2013): ""Academic Redshirting" in Kindergarten Prevalence, Patterns, and Implications," *Educational Evaluation and Policy Analysis*, 35(3), 283–297.

BEDARD, K., AND E. DHUEY (2006): "The persistence of early childhood maturity: International evidence of long-run age effects," *The Quarterly Journal of Economics*, 121(4), 1437–1472.

BERTANHA, M., AND G. W. IMBENS (2014): "External Validity in Fuzzy Regression Discontinuity Designs," Discussion paper, National Bureau of Economic Research.

BLACK, S. E., P. J. DEVEREUX, AND K. G. SALVANES (2011): "Too young to leave the nest? The effects of school starting age," *The Review of Economics and Statistics*, 93(2), 455–467.

BLAKE, P. R., M. PIOVESAN, N. MONTINARI, F. WARNEKEN, AND F. GINO (2014): "Prosocial norms in the classroom: The role of self-regulation in following norms of giving," *Journal of Economic Behavior & Organization*.

BROWNING, M., AND E. HEINESEN (2007): "Class Size, Teacher Hours and Educational Attainment*," *The Scandinavian Journal of Economics*, 109(2), 415–438.

BUCKLES, K. S., AND D. M. HUNGERMAN (2013): "Season of Birth and Later Outcomes: Old Questions, New Answers," *The Review of Economics and Statistics*, 95(3), 711–724.

CASCIO, E., AND D. W. SCHANZENBACH (forthcoming): "First in the class? Age and the education production function," *Education Finance and Policy*.

CRAWFORD, C., L. DEARDEN, AND C. MEGHIR (2007): "When you are born matters: the impact of date of birth on child cognitive outcomes in England," Open Access publications from University College London http://discovery.ucl.ac.u, University College London.

CUNHA, F., J. J. HECKMAN, L. LOCHNER, AND D. V. MASTEROV (2006): "Interpreting the evidence on life cycle skill formation," *Handbook of the Economics of Education*, 1, 697–812.

DATTA GUPTA, N., AND M. SIMONSEN (2010): "Non-cognitive child outcomes and universal high quality child care," *Journal of Public Economics*, 94(1), 30–43.

DEMING, D., AND S. DYNARSKI (2008): "The Lengthening of Childhood," *The Journal of Economic Perspectives*, 22(3), 71–92.

DUCKWORTH, A. L., AND S. M. CARLSON (2013): "Self-regulation and school success," *Self-regulation and autonomy: Social and developmental dimensions of human conduct*, 40, 208.

ELDER, T. E. (2010): "The importance of relative standards in ADHD diagnoses: Evidence based on exact birth dates," *Journal of Health Economics*, 29(5), 641–656.

ELDER, T. E., AND D. H. LUBOTSKY (2009): "Kindergarten entrance age and children's achievement impacts of state policies, family background, and peers," *Journal of Human Resources*, 44(3), 641–683.

FREDRIKSSON, P., AND B. ÖCKERT (2013): "Life-cycle Effects of Age at School Start," *The Economic Journal*.

GOODMAN, R. (1997): "The Strengths and Difficulties Questionnaire: a research note," *Journal of child psychology and psychiatry*, 38(5), 581–586.

GOODMAN, R., AND S. SCOTT (1999): "Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: is small beautiful?," *Journal of abnormal child psychology*, 27(1), 17–24.

HECKMAN, J. J., AND T. KAUTZ (2012): "Hard evidence on soft skills," *Labour economics*, 19(4), 451–464.

IMBENS, G. W., AND J. D. ANGRIST (1994): “No Title,” *Econometrica*, 62(2), 467–475.

JACOBSEN, T. N., E. A. NOHR, AND M. FRYDENBERG (2010): “Selection by socioeconomic factors into the Danish National Birth Cohort,” *European journal of epidemiology*, 25(5), 349–355.

LANDERSØ, R., H. S. NIELSEN, AND M. SIMONSEN (2013): “School Starting Age and Crime,” Economics Working Papers 2013-03, School of Economics and Management, University of Aarhus.

LEE, D. S., AND T. LEMIEUX (2010): “Regression Discontinuity Designs in Economics,” *The Journal of Economic Literature*, 48(2), 281–355.

LEUVEN, E., M. LINDAHL, H. OOSTERBEEK, AND D. WEBBINK (2010): “Expanding schooling opportunities for 4-year-olds,” *Economics of Education Review*, 29(3), 319–328.

MCCRARY, J. (2008): “Manipulation of the running variable in the regression discontinuity design: A density test,” *Journal of Econometrics*, 142(2), 698–714.

MCEWAN, P. J., AND J. S. SHAPIRO (2008): “The benefits of delayed primary school enrollment discontinuity estimates using exact birth dates,” *Journal of human Resources*, 43(1), 1–29.

MÜHLENWEG, A., D. BLOMEYER, H. STICHNOTH, AND M. LAUCHT (2012): “Effects of age at school entry (ASE) on the development of non-cognitive skills: Evidence from psychometric data,” *Economics of Education Review*, 31(3), 68–76.

OLSEN, J., M. MELBYE, S. F. OLSEN, T. I. SØRENSEN, P. AABY, A.-M. N. ANDERSEN, D. TAXBØL, K. D. HANSEN, M. JUHL, T. B. SCHOW, ET AL. (2001): “The Danish National Birth Cohort-its background, structure and aim,” *Scandinavian journal of public health*, 29(4), 300–307.

PUHANI, P. A., AND A. M. WEBER (2005): “Does the Early Bird Catch the Worm? Instrumental Variable Estimates of Educational Effects of Age of School Entry in Germany,” Publications of Darmstadt Technical University, Institute of Economics (VWL) 25840, Darmstadt Technical University, Department of Business Administration, Economics and Law, Institute of Economics (VWL).

STONE, L. L., R. OTTEN, R. C. ENGELS, A. A. VERMULST, AND J. M. JANSSENS (2010): "Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4-to 12-year-olds: a review," *Clinical child and family psychology review*, 13(3), 254–274.

THE BOSTON GLOBE (2014): "Holding kids back for kindergarten doesn't help," *Evan Horowitz*.

THE DANISH MINISTRY OF EDUCATION (2009): "Fælles Mål 2009 Børnehaveklassen," Discussion Paper 25, Undervisningsministeriets håndbogsserie.

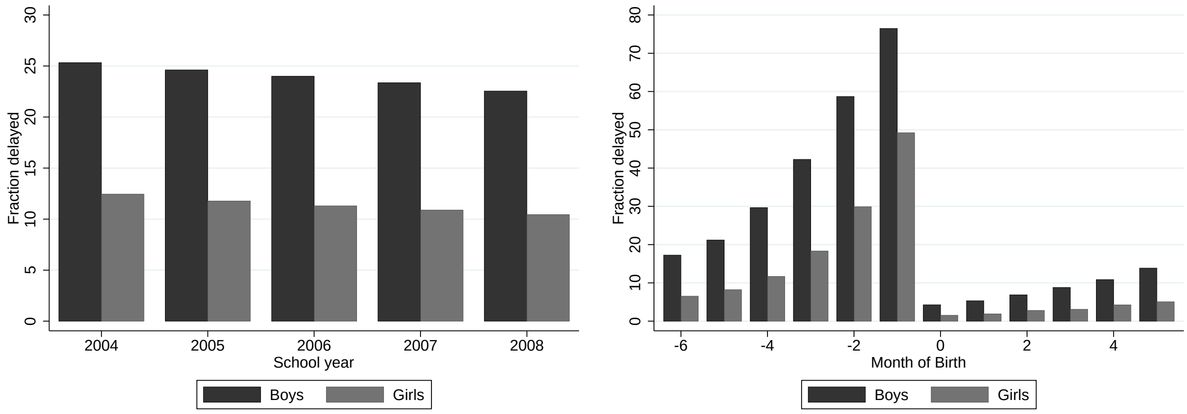
THE NEW YORK TIMES (2010): "The Littlest Redshirts Sit Out Kindergarten," *Pamela Paul*.

THE NEW YORKER (2013): "Youngest Kid, Smartest Kid?," *The New Yorker, Maria Konnikova*.

VYGOTSKY, L. S. (1978): "The role of play in development," *Mind in society*, pp. 92–104.

WHITEBREAD, D. (2011): *Developmental psychology and early childhood education*. London,Sage.

Appendices



(a) *By year of enrollment* **(b)** *By month of birth*

Figure A.1: *Share of school entrants that are delayed. Imputed by when they participated in the first National Test.*

Table A.1: *Variable descriptives, Survey sample compared to population data.*

	Population data			Survey			P-value
	Mean	SD	N	Mean	SD	N	
Years of schooling, highest among parents	14.37	2.74	481,255	15.66	2.00	36,444	0.00
Parents gross income	664.92	491.21	481,253	745.72	479.00	36,444	0.00
Mother's age when child was born	25.84	11.29	557,688	30.74	4.49	36,521	0.00
Birthweight (gr.)	3493.11	613.82	478,586	3572.34	584.82	36,141	0.00
Female	0.49	0.50	478,586	0.50	0.50	36,141	0.00
5min APGAR score	8.48	3.47	557,369	9.79	1.09	36,507	0.00

Notes: Birth weight is measured in grams. Educational length is measured in years. Parents are defined as non-western if they are immigrants to Denmark from a non-western country according to the classification by Statistics Denmark. The mother's single status is one if the child is living with the mother, and the mother is not married or cohabiting. The gross income is measured in 1,000 DKK and adjusted to the 2010 level using the consumer price index. The parents' employment is for November in the lagged year.

Table A.2: *Auxiliary RD estimates, balancing of the covariates.*

	(1)	(2)
Non-western origin	-0.00	-0.00
	(0.01)	(0.00)
Years of schooling, highest among parents	0.18*	0.09
	(0.09)	(0.06)
Parents gross income	8.10	14.71
	(20.53)	(14.71)
Mother's age when child was born	0.20	0.08
	(0.21)	(0.12)
Birthweight (gr.)	4.88	15.38
	(29.04)	(16.76)
Female	-0.01	0.01
	(0.02)	(0.01)
5min APGAR below 7	0.00	0.00
	(0.00)	(0.00)
Bandwidth	30 days	Full
Linear trend × cutoff	✓	✓
Quadratic trend × cutoff		✓

Robust standard errors in parenthesis. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Regressions of the covariates on the indicator for being born on January 1st or later as well as time trends. Each cell represents a regression and shows the point estimate on the indicator for being born January 1st or later.

Table A.3: *Placebo regressions with pre-treatment outcomes*

	(1)	(2)	(3)	(4)
Can keep occupied for 15min aged 18m	-0.02	-0.03	0.03	0.01
	(0.09)	(0.09)	(0.08)	(0.07)
Turns pictures right aged 18m	0.21	0.19	0.10	0.06
	(0.14)	(0.13)	(0.10)	(0.09)
Makes word sounds aged 18m	0.04	0.04	0.02	0.01
	(0.04)	(0.04)	(0.03)	(0.03)
Can walk up stairs aged 18m	0.00	0.00	-0.01	0.01
	(0.03)	(0.03)	(0.03)	(0.02)
Can bring things aged 18m	0.00	-0.00	0.00	-0.01
	(0.04)	(0.03)	(0.03)	(0.02)
Observations	5,816	5,816	40,749	40,749
Bandwidth	30 days	30 days	Full	Full
Covariates		✓		✓
Linear trend × cutoff	✓	✓	✓	✓
Quadratic trend × cutoff			✓	✓

Robust standard errors in parenthesis. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Covariates included are birth weight, 5 minute APGAR score, parental education, parents' age, parental income, parental employment, age at test (monthly indicators), and birth year fixed effects.