School Starting Age and Non-Cognitive Skills†

Thomas S. Dee* & Hans Henrik Sievertsen†

Version: April 30, 2015
Preliminary draft, please do not cite or circulate without authors' permission.

Abstract

An increasing number of parents, particularly those with socioeconomic advantages, have chosen to delay the school starting age of their children (i.e., academic "red-shirting"). However, a growing body of carefully identified empirical evidence provides mixed evidence on the conjectured human-capital benefits of such delays with much of the seeming benefits due merely to a delayed age at which tests are taken or to the incapacitation of older youth from risky behaviors. 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 measures of non-cognitive skills measured contemporaneously, with regard to age, for young students who may be in different grades. This study identifies the causal effects of delayed school enrollment using a regression discontinuity design based on exact dates of birth and the fact that children typically enroll in school during the calendar year in which they turn six. We find that a delayed school start dramatically reduces hyperactivity at ages 7 and 11, a measure with strong negative links to student achievement. However, the estimated effects on non-cognitive dimensions with weaker links to student achievement (emotion, conduct, peer relations, and social skills) are small and less persistent.

1 Introduction

Delaying school entry of their children – also known a academic "red-shirting" – is becoming increasingly prevalent, particularly among socioeconomic advantaged parents.[1] According to the U.S. National Center for Education Statistics six percent of all school entrants in fall 2010 were delayed, and data from Statistics Denmark reveals that in Denmark more than one out of five boys and one out of ten girls have a delayed school start.[2] However, the gains of postponing school enrollment to boost the child's development have recently been questioned both in research and in the public debate (The New Yorker, 2013). While older school starting age lowers the propensity to commit crime (Landersø et al.,

*Stanford University, Graduate School of Education & NBER. Contact : tdee@stanford.edu.

†The Danish National Centre for Social Research (SFI). Contact: hhs@sfi.dk.

†*We thank Paul Bingley, Björn Öckert, Kjell Salvanes, and seminar participants at SFI and the Copenhagen Education Network for helpful comments and suggestions. This paper uses data from the Danish National Birth Cohort (DNBC). The Danish National Research Foundation has established the Danish Epidemiology Science Centre that initiated and created the DNBC. The DNBC is a result of a major grant from this Foundation, as well as grants from the Pharmacy Foundation, the Egmont Foundation, the March of Dimes Birth Defects Foundation, the Augustinus Foundation, and the Health Foundation Sievertsen acknowledges support from The Danish Council for Strategic Research, Grant DSF-09-070295.*

[1] 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 20% group according to socioeconomic status, while 24% were children of parents in the highest 20% socioeconomic status group. Socioeconomic status is measured 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 enrollment as the year the child enters kindergarten, which in Danish is referred to as grade zero or "Børnehaveklasse".

2013) and the risk of teenage pregnancy (Black et al., 2011), both effects are driven by incapacitation rather than human capital effects. Moreover, the positive effects of school starting age on tests in primary school (Bedard and Dhuey, 2006) seem to be dominated by age-at-test effects, as Black et al. (2011) show using post schooling test scores. Research on the causal evidence on a direct effect of delaying school entry on children's skills suffers from the fact that the link between school starting age and primary school test scores is "fundamentally unanswerable" (Black et al., 2011), as it is empirically impossible to control for the fact that children who are older at enrollment also are older when tested within school.

Using Danish data we provide evidence on the causal effect of school starting age on children's human capital, by using out-of-school measures of non-cognitive skills and holding age constant. We focus on non-cognitive skills because these skills (1) are less dependent on grade (which has to vary, when age is held constant), (2) are important for the children's ability to acquire cognitive skills, and (3) are important for later life outcomes in terms of crime, health, and the labor market (Cunha et al., 2006). The study most similar to ours is by Elder and Lubotsky (2009), who show that older school starting age causes a reduced likelihood of an ADHD diagnosis, but these effects may be driven by a relative age effect.

Denmark provides an excellent case for evaluating the direct effects of school starting age for three reasons: One, the universal day care system and the centrally specified school program constitute homogeneous control and treatment environments.[3] Two, the existence of a discontinuity in expected school starting age created by the rule that Danish children are supposed to enter school the calendar year they turn six. Three, the availability of register-based data linked to a large survey on 55,000 children's skills measured in terms of the Strength and Difficulties Questionnaire (SDQ) (Goodman, 1997). The SDQ provides information on children's non-cognitive skills on five dimensions: conduct, emotional symptoms, hyperactivity, peer problems, and pro-social behavior.

Using a fuzzy regression discontinuity design we find that older school starting age causes better non-cognitive skills at age seven and at age 11 measured by the aggregated SDQ score. The effects are driven by better outcomes on the hyperactivity scale of the SDQ, while the other dimensions are less affected. Linking the SDQ scores to the children's later performance in tests of cognitive skills we find that especially the hyperactivity scale is closely related to the test performance in mathematics and reading. Assessing the heterogeneity in the treatment effects by means of marginal treatment effects reveals that the average effect is driven by girls with a low level of latent ability.

This paper proceeds as follows: Section two provides a brief review of the existing literature and a description of the institutional setting and the SDQ which is the outcome variable used in this study. Section three presents the empirical strategy. Section four describes the applied data. Section five presents the results. Section six concludes this paper.

2 Background

2.1 School Starting Age, Years of Schooling and Age at Test.

Several studies have attempted to identify the short-run effect of school starting age on the child's development. Bedard and Dhuey (2006) consider a sample of twenty countries and find that being older at school enrollment causes better fourth and eight grade test scores in mathematics and science, with the exception of children in Denmark and Finland. Their identification is based on predicted school starting age using country specific cutoff dates and individual birth dates. To check for potential endogeneity due to to season of birth effects they also run a pooled regression using variation between

[3] According to Statistics Denmark more than 95 percent of a cohort is 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.

countries. However, as all other studies on the short-run effect of school starting age, Bedard and Dhuey (2006) faces the challenge of the following linear relationship between school starting age (SSA), age-at-test (AAT) and years of schooling (YOS):

$$SSA = AAT - YOS \tag{1}$$

In samples of children with variation in school starting age (SSA), there must also be variation in either age at which we measure the outcome (AAT) or the years of schooling (YOS). Most studies keep YOS constant and thus identify the combined effect of an older school starting age and being older when tested. In other words, as children who start later will always be older when they are tested within school, the effect of school starting age on test scores may be due to the age at the time of the test, and not because of a direct effect of school starting age. The positive relationship between school starting age and test scores found by among others Bedard and Dhuey (2006) may be purely driven by the fact that children who enroll later also are older when tested. Focusing on non-cognitive skills we take a different approach and keep age-at-test constant. We identify the combined effect of starting school later and having received *less* schooling. Under the assumption that non-cognitive skills are not decreasing in years of schooling, a positive relationship between school starting age and non-cognitive skills can be interpreted as a (lower-bound of) the effect of school starting age. We are only aware of one other study attempting to exploit out-of-school measures of skills, holding age constant: Mühlenweg et al. (2012) use a sample of 360 children from the German Rhine-Neckar region to show that later school starting age is related to being less hyperactive at age eight. They exploit the panel structure of their data to compensate for the issue that variation in school starting age is only obtained through variation in birth date as they have no variation in cutoff dates.

Black et al. (2011) use Norwegian data in a regression discontinuity framework and to test the effect of school age on post compulsory schooling outcomes. They can therefore control for both school starting age and age-at-test. They find that the effects of being older at school enrollment on test-scores primarily is driven by the age-at-test effect. They find that starting school later has a positive effect on mental health at age 18 for boys and a negative effect on the likelihood of teenage pregnancy for girls. However, they also find that older school starting age increases the likelihood of pregnancy within 12 years after school enrollment. The negative effects on teenage pregnancy are therefore driven by the fact that children who start later also are older when they leave school, leaving them less time to participate in risky behavior as teenagers.

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 increases educational attainment. Their sample period includes a school reform which postponed tracking, and their results show that in the period with postponed tracking the effects of school starting age are smaller. While they find that the effects on discounted life-time earnings on average are very, they also find positive earnings effects of school starting age for individuals with low-educated parents.

Landersø et al. (2013) use exact day of birth - in contrast to Black et al. (2011) and Fredriksson and Öckert (2013) who use month of birth - and compare a sample of the Danish population born in January and December. Exploiting the January 1st school starting age cutoff date in Denmark, they find that being older at school enrollment lowers the propensity to commit crime. The result is however, mainly driven by an incapacitation - and not a human capital - effect.

A few studies have assessed the importance of school starting age for children's behavior in the short run. Elder and Lubotsky (2009) exploit variation in school starting age cutoff dates across U.S. States and find that being one year older at school entry reduces the probability of a Attention Deficit/Hyperactivity Disorder (ADD/ADHD) diagnosis between kindergarten and fifth grade. Exploiting variation in school entry cutoff dates in a pooled sample of 17 countries, Mühlenweg (2010) shows that older school starting age is linked to a reduced likelihood of being "victimized" in school. The effects are driven by a school starting age effect or age-at-test effects.

This present paper makes four contributions to the literature: One, we employ a standardized measure of non-cognitive skills in terms of the SDQ. Two, our outcomes are measured out of school, which reduces the problem that measured differences in non-cognitive skills may be caused by relative age effects.[4] Three, our outcome measures are independent of grade, holding age constant, allowing us to rule out age-at-test effects. Four, we exploit repeated measures of the same outcome variables for the same children, at both age seven and age 11.

2.2 Institutional setting

Daycare in Denmark is almost exclusively publicly provided and organized by the municipality. Child care consists of center based nurseries for children aged up to two and daycare for children aged three to six. In addition to the center based nurseries, municipalities provide family day care. Requirements to center based day care 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. Requirements to family day care are lower.

Compulsory schooling begins in "grade zero" (also called kindergarten class) in August the year the child turns six. Until 2009 grade zero was not mandatory, but 98% of a cohort attended grade zero (Browning and Heinesen, 2007). Compulsory schooling ends after ten years of schooling or in August of the year the child turns 17. Figure 1 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 allows access higher education.

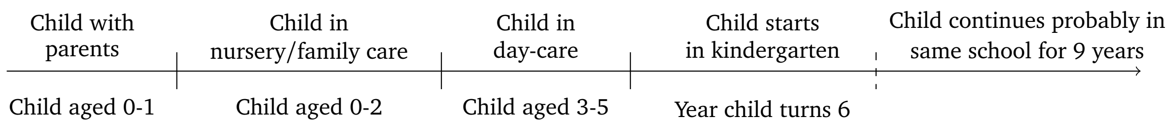


Figure 1: *Timing of childhood*

As children are supposed to enroll in school the year they turn six, school starting age increases discontinuously from December 31 to January 1. To illustrate this we compare the events in Figure 1 for a child born December 31 to a child born January 1 in Table 1. Children who comply to the rules will be one year older at school enrollment if they are born on January 1 compared to if they are born one day earlier. It is possible to postpone enrollment in school. This requires considerable effort of the parents and involves meeting representatives from the future school. Based on individual evaluations children may enroll in grade zero one year earlier, if their birthday is before October 1.

Kindergarten class is part of the primary school and free of charge in the public schools. 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).[5] The formation of the class is based on either pedagogical or practical considerations, and the principle is the same as for 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).

[4]Elder and Lubotsky (2009) for instance argue that the identified effects on ADHD may be driven by the facts that older children may be under-diagnosed, because they are compared to their younger classmates.

[5]The individual plans were introduced in 2006.

Table 1: *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 daycare	Months 37-65	Months 37-77
Enroll in grade zero	Month 66	Month 78

As almost all children attend daycare before they enroll in school, the control environment is very homogeneous. Identified effects of school starting age are therefore not a result of a different pre-treatment environments. Also, the kindergarten class is centrally defined and constitutes a homogeneous treatment environment.

2.3 The Strengths and Difficulties Questionnaire (SDQ)

The SDQ is a questionnaire developed by the English child psychiatrist Robert N. Goodman in the mid 1990s. The objective of the SDQ is to provide a tool to describe behavior of children aged 4 to 16 (Goodman, 1997). Compared to two established screening devices, the Rutter questionnaire and the Child Behavior Checklist (CBCL), the SDQ is distinctive in that it is shorter (The SDQ has 25 questions in contrast to 120 in the school-age CBCL), it is uniform, and it assesses both children's strengths and difficulties.

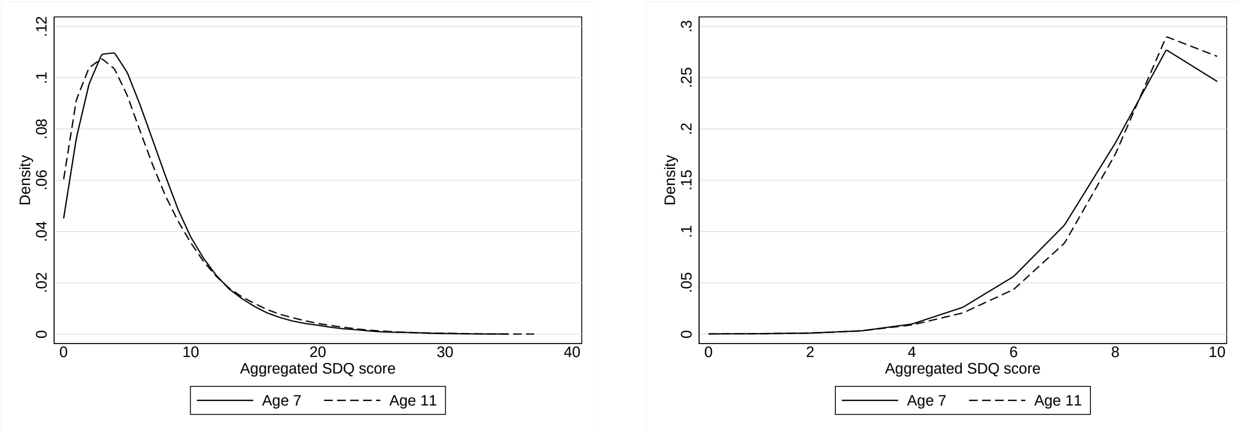
The questionnaire is filled out by a parent in our case, but can also be filled out by a teacher. For each question there are three possible answers: Not True, Somewhat True, Certainly True. An example of a question is that the child is "Restless, overactive, cannot stay still for long". The SDQ scores are computed according to a standardized aggregation procedure.[6] The questionnaire consists of 25 questions covering five dimensions of children's behavior related to non-cognitive skills:

1. Emotional Symptoms Scale
2. Conduct Problems Scale
3. Hyperactivity Scale
4. Peer Problems Scale
5. Pro-social Scale

Higher values means worse skills, with the exception of the pro-social scale. The aggregated measure (Total difficulties score) includes all dimensions except for the pro-social scale. Figure 2 shows the distribution of SDQ scores in our sample. For the first four dimensions, scores between 0 and 13 are regarded as normal, while scores 14-16 are borderline and scores above 16 are regarded as abnormal. For the pro-social scale 6-10 is normal, 5 is borderline, and 0-4 is abnormal.

Goodman (1997) finds a close correlation between the SDQ and the Rutter questionnaire using a sample of about 400 English children aged four to 16. Goodman and Scott (1999) compare the SDQ to the Child Behavior Checklist (CBCL) on about 130 English children aged four to seven from both a high and a low risk sample. The measures are highly correlated and are both able to identify the

[6]The aggregation procedure is described on the website www.sdqinfo.com. We compared the outcome of the standardized aggregation procedure to a principal component analysis (PCA) using our data. The PCA revealed the same five dimensions as the standardized procedure.



(a) *The Total Difficulties Score (includes conduct, emotional, hyperactivity, and peer problems scales. Larger values implies worse skills.)*

(b) *The Pro-social scale. Larger values implies better skills.*

Figure 2: *The distribution of SDQ scores in our sample. The SDQ is reported by the mother for the same child at age seven and age 11*

two underlying samples. In addition, interview-based evaluations seem to be stronger correlated to the SDQ than to the CBCL.

2.4 Use of the SDQ in research

Using the British Cohort Study, Meschi et al. (2008) test the importance of parents numeracy and literacy skills for the child's cognitive and non-cognitive skills, aged three to six. In the unconditional regressions there is a strong correlation between parental skills and both cognitive and non-cognitive child skills. Once they control for socio-economic background and child characteristics the correlation between parents skills and the SDQ measure of non-cognitive skills becomes insignificant.

In a Danish context Datta Gupta and Simonsen (2010) use SDQ scores computed from the Danish Longitudinal Survey of Children (DALSC) to assess the effects of center based child care versus family day care for children aged three on non-cognitive skills. They find a negative effect of being in family day care for boys of lower educated mothers. Using the same data and identification strategy, Datta Gupta and Simonsen (2012) evaluate SDQ outcomes for the children aged 11, but find no effects.

2.5 SDQ and primary school test scores

While non-cognitive skills are important in themselves, they might also be important for the development of cognitive skills. The established link between school starting age and cognitive skills has been motivated by two potential channels (Bedard and Dhuey, 2006): (1) Age-at-test effects, because children that start later are older when evaluated. (2) A maturity effect, because children who start later are more mature, and therefore more able to comprehend the human capital inputs they receive in school. If the latter is the case, we should expect to find a correlation between SDQ scores and children's performance in tests of cognitive skills.

To assess the link between cognitive and non-cognitive skills we use data on the mother reported SDQ for children aged seven and data from the mandatory tests in primary school.[7] We regress the performance in three tests in Reading and two tests in Mathematics on the five dimensions of non-cognitive skills in the SDQ. In each regression we include school fixed effects to handle the fact that

[7]The data is described in section 4.

Table 2: *Test scores in Danish and mathematics and the five dimensions of the SDQ.*

Subject Grade	- - - Reading - - - 2	 4	 6	- - - Mathematics - - - 3	 6
Emotional Symptoms 7	0.03***	0.04***	0.03***	0.00	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Conduct Problems 7	-0.05***	-0.06***	-0.06***	-0.05***	-0.07***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Hyperactivity 7	-0.16***	-0.16***	-0.15***	-0.16***	-0.14***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Peer Problems 7	-0.01	-0.01	0.00	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Pro-social Behavior 7	-0.05***	-0.05***	-0.04***	-0.04***	-0.04***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
N	20371.00	37209.00	24019.00	31214.00	23989.00

Standard errors clustered on the school level in parenthesis. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Each column shows results from one regression with test scores as the dependent variable, the five SDQ dimensions as independent variables and a set of covariates. Covariates included are birth weight, 5 minute APGAR score, parental education, parents' age, parental income, parental employment, mother's civil status, age at test monthly indicators (both for SDQ and the mathematics/Danish tests), test year, school, and birth year fixed effects. Both the five SDQ scores and the test scores are standardized.

schools decide how the tests are carried out. We also control for age-at-test, both for the SDQ scores and for the tests by means of monthly indicators.[8] Table 2 shows the coefficients on each of the five dimensions for the five separate regressions.

The sample consists of children born between 1998 and 2003 who attended a test in the years 2010 to 2013. The samples differ across columns in Table 2. The first column consists of the sample for which second grade test results are available, which roughly covers birth cohorts 2001 to 2003. The fifth column includes children for whom sixth grade test results are available, which corresponds to the cohorts 1998 to 2000. Considering that the samples, grades and subjects differ across the five columns the correlations between the SDQ scores and the test scores are remarkably constant. A one standard deviation better (i.e. lower) conduct score is associated with a 0.05-0.06 standard deviation higher test score in Reading and Mathematics in grades two to six. The correlation between the hyperactivity scale and the test results is almost three times the coefficients on the conduct scale: A one standard deviation better hyperactivity score is linked to 0.14-0.16 standard deviation better test scores. Peer problems are almost unrelated to cognitive skills, while better pro-social and emotional skills are related to worse test performances. The correlations are almost identical across gender as Tables A.1 and A.2 in the appendix show.

Using data from the Danish Longitudinal Survey of Children (DALSC), Table A.3 in the Appendix shows the correlation between SDQ scores and the brief version of the Big Five Personality Traits (Rammstedt and John, 2007) and a brief version of Raven's Progressive Matrices. The DALSC is a survey of about 5,000 children born in the fall of 1995. The parents and the children have been surveyed several times. The data used here is from the interviews conducted in 2011. The columns (1) to (6) show that the mother reported SDQ is closely related to the Big Five and the Raven score. In-

[8]While the tests are standardized and computer based, the school decides whether the tests are open book or not.

terestingly the SDQ's hyperactivity dimension is closest related to the Raven score, a measure of the individuals "reasoning ability".

Intuitively it makes sense that the non-cognitive skills that are most closely linked to cognitive skills are hyperactivity and conduct. Hyperactivity probably gives a good indication for whether the child can sit still and concentrate in class, while conduct gives an indication of whether the child can behave and follow instructions by a teacher. It is therefore of special interest whether school starting age affects these two dimensions, as this could give us some indication of whether the effect of school starting age on cognitive skills also is a human capital effect, or whether the entire effect is driven by age-at-test effects. The finding that correlations are very constant across samples, subjects, and grades indicates that the SDQ measures capture important aspects of children's non-cognitive skills.

3 Empirical framework

3.1 Identifying strategy

Formally we assume that the relationship between school starting age (SSA) and non-cognitive skills (Y) for individual i with covariates $\boldsymbol{X_i}$ can be represented by the following linear relationship:

$$Y_i = \beta_0 + \beta_1 SSA_i + \phi \boldsymbol{X_i} + e_i \tag{2}$$

Identifying the causal effect of school starting age on non-cognitive skills is challenging because children with less developed non-cognitive skills are more likely to enroll later. The ordinary least squares (OLS) estimates of β_1 in equation (2) are therefore potentially downward biased. For example if children have difficulties concentrating and focusing, one might decide that starting school is too early. Consequently, children with older school starting age will have worse non-cognitive skills than children who start when they are younger. However, the causal effect of school starting age on non-cognitive skills may still be positive.

To identify the causal effect we need an instrument Z, which is unrelated to unobserved characteristics of the child that affect the outcome and is related to school starting age. In our case we exploit that in Denmark children are supposed to enroll in school the year they turn six. We create an instrument that takes the value of one if the child is born January 1st or later for a year running from July 1st to June 30th, in a fuzzy regression discontinuity design. The fuzziness is created by the fact that parents and their children not necessarily comply with the January 1st cutoff rule.

While evidence shows that season of birth is not random with respect to parental characteristics (Buckles and Hungerman, 2013), it is unlikely that the exact date of birth is related to individual observed and unobserved characteristics. In practice two approaches can be used to handle this issue. One, by considering only the local sample around the January 1st cutoff in a non-parametric approach, or two, by considering the full sample and specifying the relationship between season of birth and age at school enrollment parametrically. In both cases the first stage regression is specified as follows:

$$SSA_i = \gamma_0 + \gamma_1 after_i + g(days_i) + after_i \times g(days_i) + \boldsymbol{\rho X_i} + \epsilon_i \tag{3}$$

Where *after* is an indicator for whether the individual was born between January 1st and June 30th. We center the forcing variable, birthday, to January 1, so that the year runs from July to June. The variable *days* is the number of days from January 1, $\boldsymbol{X}$ is a vector of controls including parents income and education, and ε is random noise. The function $g(\cdot)$ is a polynomial function of day of birth, which is included to control for trending behavior that is continuous. The predicted exogenous variation in school starting age (*SSA*) from this first stage regression is then included in the second stage regression of the following equation:

$$outcome_i = \beta_0 + \beta_1 \widehat{SSA_i} + g(days_i) + after_i \times g(days_i) + \phi \boldsymbol{X_i} + e_i \tag{4}$$

Where the coefficient β_1 identifies the causal effect of school starting age on the outcome variable. For the local specification we consider a 30 day bandwidth with linear trends interacted with the cutoff date as in Landersø et al. (2013). We assess the robustness of our results to the choice of bandwidth. For the full sample analysis, using a July to June sample, we select the polynomial function based on a graphical judgment and by comparing the Akaike Information Criteria (AIC) for various specifications as suggested by Lee and Lemieux (2010).

3.2 The LATE interpretation

The coefficient β_1 captures the constant effect of starting school later for the population if the treatment effect is homogeneous. As children may differ substantially in their school readiness it is likely that the treatment effect is not constant across the population. The identified treatment effect is therefore only an average treatment effect. Because we exploit exogenous variation in school starting age caused by the rule that children should start school the year they turn six, we only have exogenous variation in treatment for those who comply with these rules. In this case the identified treatment effect is a local average treatment effect (LATE). Formally the treatment parameter can be expressed in terms of the Wald estimator (Angrist and Pischke, 2008):

$$\beta_1^{LATE} = \frac{E[Y|Jan] - E[Y|Dec]}{P(SSO|Jan) - P(SSO|Dec)} \tag{5}$$

Where the numerator is the reduced form relationship between date of birth and the outcome Y and the denominator the first stage relationship between date of birth and being older at school enrollment. Note that for always takers (those who always start school late) and never takers (those who never start school late) the expression is not defined as the denominator is zero. β_1 can be interpreted as the LATE effect if the instrument is valid and it increases the propensity of being older at school enrollment monotonically. In other words, we interpret our estimates as LATE effects by the assumption of having no defiers, that is children where parents always choose the opposite of what the school starting age rules recommend. It seems unlikely that parents plan to enroll their child one year too early if it is born in January, but plan to postpone enrollment one year if it is born in December.

3.3 Characterizing the compliers

While being born in January increases the expected school starting age, it is important to stress that there are some children who are unaffected by this rule. For simplicity consider school starting age as binary variable that takes the value of one if the child is older than six years and six months at school enrollment, $SSO = 1(SSA > 6.5)$. Children with very low school readiness will probably never be treated, whether they are born in December or January. Likewise, children who are very school ready will probably always be treated, whether they are born in December or January. The treatment effect is therefore only identified for the subgroup of the population that complies to the rule, i.e. changes behavior when born after December 31st.

While it is empirically impossible to identify the subpopulation who complies to the school starting rule, it is possible to characterize their observable characteristics. Formally we can calculate the probability that a complier has a characteristic $x_i = a$ using the following expression (Angrist and Pischke, 2008):

$$P[x_i = a|P(SSO|Jan) > P(SSO|Dec)] = \frac{P_i(SSO|Jan, x_i = a) - P_i(SSO|Dec, x_i = a)}{P_i(SSO|Jan) - P(SSO|Dec)} \quad (6)$$

That is, the probability that a complier has a characteristic a is given by the ratio of the first stage coefficient for individuals with this characteristic to the overall first stage coefficient.

3.4 Identifying the marginal treatment effect

If we consider school starting age as a binary variable as above, it is possible to identify the marginal treatment effect (MTE) which is the treatment effect for children at the margin of being old at school enrollment:

$$\beta_1^{MTE} = E[Y|\epsilon = \epsilon^*, \boldsymbol{X} = \boldsymbol{X}^*] \quad (7)$$

Where ϵ_i^* is the error term in equation (3) moving the individual to the margin of treatment. Individuals with a very large observed propensity of treatment will need a numerically very large error term to make them "indifferent" between treatment and non-treatment. To calculate the MTE we follow the parametric approach described by Heckman et al. (2006):

1. Estimate a probit for the propensity of treatment
2. Specify the area of common support, that is observed propensities of treatment where we observe both treated and untreated.
3. Compute the MTE as a function of u, where $\epsilon = F_u(u)$

4 Data

4.1 The data structure

We create our analysis sample by combining Danish administrative registers from Statistics Denmark with data from the Danish Ministry of Education on the compulsory tests in grades two to eight, and data from the Danish National Birth Cohort Survey (DNBC) (Olsen et al., 2001). The registers provide information on the child's birthday (the forcing variable), the test data is used to impute school starting age (the treatment), and the survey data provides information on the non-cognitive skills in terms of the parent reported SDQ scores (the outcomes). The three data sources are described below.

4.2 The administrative registers

The administrative data consists of several individual registers including the birth records, the income registers, the annual register based labor force statistics (RAS), and the education registers. All datasets are hosted by Statistics Denmark and linked by the unique personal identifier.

For the children we use information from the registers on birthday, birth weight, 5 minute APGAR score[9], and the gestational age. For the parents we use information on gross annual income, the labor market attachment in November, 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.

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

4.3 The National Test data

The National Test is an obligatory assessment of the school children's cognitive skills in grades two to eight. The tests were introduced in 2010. The tests are used as a tool for the teacher to assess the child's development, and they have no direct impact on the child's continued schooling (except that the teacher can use test results when grading the pupils). The test data provided by the Ministry of Education provides information on the exact testing time, the grade, and the subject. We use this information to impute the school starting age. For example if the child took the second grade test in spring 2010 the child must have enrolled in school in August 2007, as the school year runs from August to June.

It may be the case that children who enroll early also are more likely to retake kindergarten class, which would cause a non-classical measurement error in school starting age. We assess this issue by also running regressions using parent reported school starting age.

4.4 The Danish National Birth Cohort (DNBC)

The DNBC is a Danish nation-wide cohort study in which 101.042 pregnant women were sampled over the period 1997 to 2003. The mother was interviewed twice during pregnancy, when the child was six months old, when the child was 18 months old, when then child was seven years old, and when the child was 11 years old. All interviews contain questions on the parents' behavior (smoking, drinking, sport activities, and employment), the mother's health and the child's health, cognitive, and non-cognitive skills (at age 18 months and seven years). 92,892 mothers replied to the first interview, 66,764 replied to the fourth interview (child aged 18 months), 57,280 replied to the fifth interview (child aged seven years), and 66,710 replied to the sixth interview (child aged 11 years).

Attrition from the sample is not random. Mothers who were not in the labor force constitute a share in the survey which is about 60% lower than in the population data, single women are underrepresented by about 30%, and low educated women by about 40% (Jacobsen et al., 2010). Table A.4 in the appendix provides a comparison between the survey sample and the full population of children born in the period 1997 to 2003 for the covariates used in this paper. With the exception of APGAR score all variables have significantly different means. Children in the survey have a higher birth weight, and have parents who have a higher income and completed more schooling. As academic red-shirting is especially prevalent among socioeconomic advantaged parents, we may therefore find that parents in the survey data comply less to the school starting age rule than the parents in the population. We discuss the implications for the external validity of our results in section 6.

In this paper we only use data from the fifth and sixth interview of the DNBC, when the children are respectively approximately seven years and two months old and 11 years and three months old. In these interviews the parents answered the 25 questions of the SDQ, which we use to create the six outcome variables (the five dimensions of the SDQ and the aggregated score). The fifth interview also contains mother reported school starting age which we use to test the validity of our school starting age measure. We also use information from the fourth interview on cognitive and non-cognitive skills before school enrollment as a placebo test.

4.5 Data selection and descriptive statistics

Table 3 provides an overview of the sample selection. The DNBC is a sample of children born 1997-2003. When the child was seven years old (ie. in years 2004-2010) 55,498 parents completed the SDQ questionnaire. Four years later (ie. in years 2008-2014) 47,042 parents replied completed the SDQ questionnaire. The data thus covers approximately 10 percent of all children born in that period. We delete some observations the school starting age or parental characteristics are missing. The final sample consists of 53,856 SDQ outcomes for children aged seven years old and 44,513 SDQ outcomes for the same children four years.[10] Appendix Figures A.2 and A.3 show the distribution of the date of birth for the final sample. While we only delete very few observations, it is important for identification strategy that the deleted observations are randomly allocated around January 1st cutoffs, which is the case as show by Figures A.4 and A.5 in the Appendix. As we present results using both a local 30 day

Table 3: *Sample selection*

		SDQ7	SDQ11
All children born 1997-2003	526,418		
In DNBC and complete SDQ records		55,498	47,042
SSA is missing		- 245	-1,412
Admin. data is missing		-536	-470
Parental characteristics are are missing		-861	-647
Final sample		53,856	44,513

bandwidth and the full sample, Table 4 provides sample means, standard deviations and the number of observations for the covariates and dependent variables in these two samples. Average school starting age is 6.26 in the local sample, which is slightly higher than in the full sample. Both the covariates and the dependent variables are very similar across these groups.

Table 4: *Variable descriptives*

	30 day bandwidth			Full sample		
	Mean	SD	N	Mean	SD	N
School starting age	6.26	0.56	9,440	6.12	0.51	64,372
School starting age>6.5	0.69	0.46	9,440	0.19	0.39	64,372
Non-western origin	0.02	0.13	9,219	0.02	0.12	63,325
Parents' years of schooling	15.34	1.98	9,219	15.40	1.96	63,325
Parents' gross income	683.40	306.99	9,219	685.46	364.42	63,325
Mother's age when child was born	30.68	4.27	9,219	30.63	4.19	63,325
Father's age when child was born	32.96	5.24	9,219	32.85	5.10	63,325
Birthweight (gr.)	3533.32	596.40	9,327	3558.77	592.61	63,716
Female	0.50	0.50	9,327	0.49	0.50	63,716
5min APGAR below 7	0.01	0.08	9,327	0.01	0.08	63,716
Missing parents covariate data	0.02	0.15	9,440	0.02	0.13	64,372
Missing admin. data	0.01	0.11	9,440	0.01	0.10	64,372

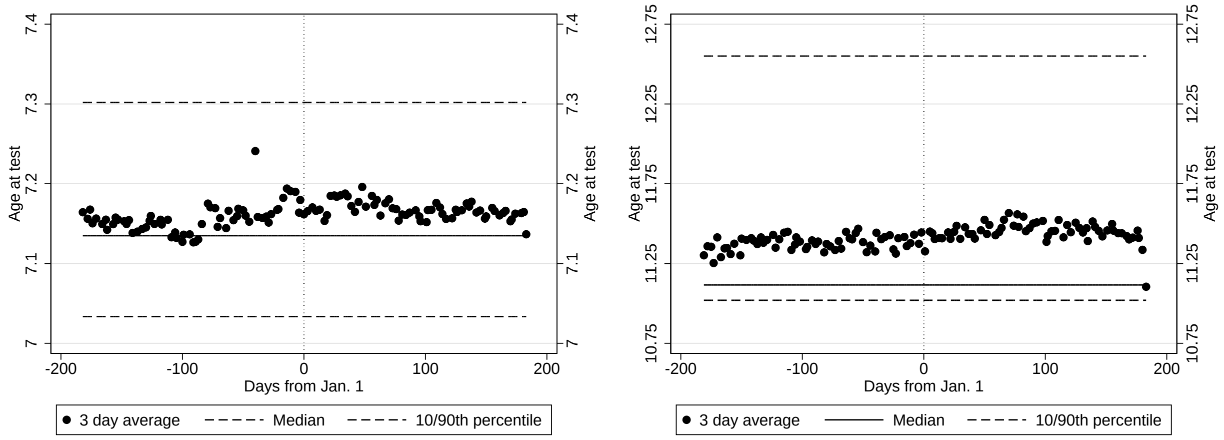
The 30 day bandwidth includes all children born within 30 days of January 1st. Birth weight is measured in grams. Educational length is measured in years. Parents are defined as non-western if they are immigrants to Denmark from a non-western country according to the classification by Statistics Denmark. 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 relative to the child's birth year.

4.6 Age at test

A key advantage of our data is that the outcome variables are independent of grade, so that all children have the same age when they are measured, independently of their enrollment age. This is confirmed by Figure 3, which shows the average age at SDQ measurement in three day bins, as well as the median and

[10]While most children in our sample is observed both at age seven and age 11 there are children who are only observed at either age seven or age 11.

the 10th and 90th percentiles. There seems to be small seasonality effects for the age seven interview, but there is no sign of a jump around January 1st. Also note that 80 percent of all children were between 7 and 7.3 years old at the measurement of non-cognitive skills. At age 11 the seasonality pattern is weaker. As for the age seven interview, there is no jump at January 1st.



(a) *Age 7* **(b)** *Age 11*

Figure 3: *The child's age in years at the time the parent answers to the 25 elements of the SDQ in the DNBC questionnaire. Each dot shows the mean age for a three-day bin.*

It is important that children on average are 7.1 and not exactly 7 (and age 11.3), because this implies that parents to children born around January 1st are interviewed in February. If the interview was exactly seven years after birth, jumps in the SDQ could be due to Christmas.

5 Results

5.1 Validity of the RD design

Before we assess the results of our analysis we first carry out a number of tests of whether the setting provides a valid RD design. Our instrument, the indicator variable for whether the child was born after January 1st, has to be uncorrelated with unobserved characteristics affecting the child's skills. Parents that consider the impact of date of birth on school enrollment six years later, are probably not a random sample of parents. Rather, such parents are a selected subsample of parents who are more prepared and potentially offer more support for their child. If parents can manipulate the date of birth perfectly the children born just after January 1st will probably have systematically different unobserved characteristics to those born just before January 1st. In this section we assess whether this is the case in three ways: (1) We evaluate the distribution of births over the cutoff. We would expect an increase in births just after January 1st, if parents have perfect control over the forcing variable. (2) We compare the mean values for observable characteristics just before and after the cutoff, to assess whether children born after the cutoff have better observed characteristics than would be expected if the parents manipulate the day of birth. (3) We graphically as well as in a regression framework test for any jumps in observed characteristics over the cutoff date.

Figure 4 shows the distribution of date of birth in our sample. We observe no spike around January 1st. There is a small drop in births at December 31st and January 1st, but the low level is not different from the drop one week earlier. The pattern is probably driven by a Christmas and New Years effect, and not that parents manipulate the date of birth. This effect may also be problematic, if a certain

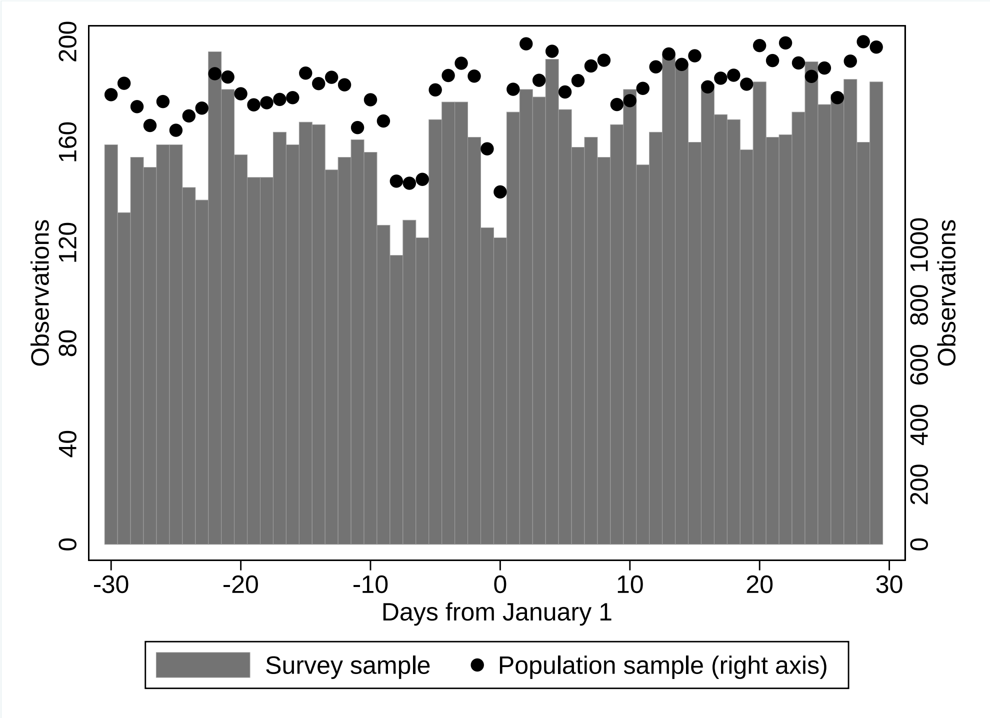


Figure 4: *Observations by date of birth, survey data and population data. The survey data is the data used in our analysis, and the population includes all children born in Denmark in the period 1998-2003.*

group of the population is affected by these events, in which case we would expect to find differences in observable characteristics across the cutoff.

Table 5 shows the sample means for the covariates included in the regression, 30 days before and after January 1st. The rightmost column presents the p-value from a t-test on the equality of the means in the two groups. The first row shows that school starting age are significantly higher for children born in January. Also, as shown in the second row, 80 percent of all children born in January are at least 6 years and 6 months at school enrollment, while this proportion is only 57 percent for December children. None of the child and parent characteristics are statistically different in January compared to December.

For the outcome variables in the lower part of Table 5 it is seen that both at age seven and age 11 the Total Difficulties Score is significantly lower for the children born in January compared to those born in December. At both ages this seems to be driven by significantly lower Hyperactivity scores.

Sample means may conceal jumps in covariates at the cutoff, as any trending behavior is ignored. We therefore use the same specification as in the first stage regressions (equation (3) without covariates) and include each covariate as a dependent variable. Table A.9 in the appendix shows the coefficient on the indicator for being born after January 1st for each of the covariates. None of the the covariates show signs of jumps at the cutoffs in neither the 30 day local specification nor in the full sample parametric specification. An alternative testing strategy is to regress the outcome variable on all covariates (without trends) and compute the predicted values (the y-hat). The y-hat then represents the average of the covariates, weighted by their influence on the dependent variable. In Table 6 we show the outcome of regressing this weighted average on the cutoff and time trends for for each of the 12 dependent variables. As for the single-covariate regressions, there is no sign of a jump in any of these specifications.[11]

[11]Note that both Table A.9 and 6 show uncorrected standard errors and significance levels. Any corrections for multiple

Table 5: *Comparison of means, Sample: 30 days before and after January 1st*

	30 days before			30 days after			P-val
	Mean	SD	N	Mean	SD	N	
School starting age	6.16	0.61	4,449	6.35	0.49	4,991	0.00
School starting age>6.5	0.57	0.50	4,449	0.80	0.40	4,991	0.00
Non-western origin	0.02	0.13	4,278	0.02	0.13	4,941	0.61
Parents' years of schooling	15.32	1.94	4,278	15.37	2.01	4,941	0.25
Parents' gross income	684.87	327.04	4,278	682.12	288.55	4,941	0.67
Mother's age when child was born	30.68	4.27	4,278	30.69	4.27	4,941	0.93
Father's age when child was born	32.92	5.19	4,278	32.99	5.29	4,941	0.53
Birthweight (gr.)	3523.30	612.04	4,397	3542.26	582.02	4,930	0.13
Female	0.49	0.50	4,397	0.50	0.50	4,930	0.49
5min APGAR below 7	0.01	0.08	4,397	0.01	0.08	4,930	0.98
Total Difficulties 7	0.04	1.00	3,679	-0.04	0.97	4,068	0.00
Emotional Symptoms 7	0.05	1.01	3,679	0.01	0.99	4,068	0.07
Conduct Problems 7	0.01	0.99	3,679	-0.02	0.98	4,068	0.23
Hyperactivity 7	0.02	1.01	3,679	-0.08	0.97	4,068	0.00
Peer Problems 7	0.02	1.02	3,679	0.01	0.99	4,068	0.39
Pro-social Behavior 7	-0.01	1.00	3,662	0.04	0.98	4,055	0.04
Total Difficulties 11	0.05	1.01	3,219	-0.05	0.97	3,593	0.00
Emotional Symptoms 11	0.02	1.01	3,219	-0.01	0.99	3,593	0.16
Conduct Problems 11	0.02	1.02	3,219	-0.02	0.98	3,593	0.09
Hyperactivity 11	0.05	1.02	3,219	-0.08	0.96	3,593	0.00
Peer Problems 11	0.05	1.02	3,219	-0.01	0.99	3,593	0.02
Pro-social Behavior 11	-0.01	1.00	3,219	0.02	1.00	3,593	0.27

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. The rightmost column presents the p-value from a t-test on the equality of the means in the two groups, assuming equal variances.

Figures A.6 and A.7 in the Appendix show the mean values in three day bins for all the covariates over the full sample period. There are no jumps on the the covariates around the January 1st cutoff and thus there is no indication of manipulation of the forcing variable.

5.2 Model selection

We now turn to the selection of the parametric specification. A good starting point for selecting the parametric specification is a graphical inspection of the reduced form relationships between the outcome variables and the forcing variable, the date of birth. Figures A.8 and A.9 in the Appendix show this relationship using a three day bin with a quadratic fit on each side of the cutoff. For all outcomes this quadratic specification captures the development in outcome scores reasonably well.

Table 7 provides a more rigorous test of the parametric specification by comparing the Akaike Information Criterion (AIC) across three types of parametric specifications. The comparison does not

testing will make the conclusions of no correlation even stronger.

Table 6: *Auxiliary RD estimates, balancing of the covariates.*

	(1)	(2)
Yhat (Total Difficulties Score, age 7)	-0.01	-0.01
	(0.01)	(0.01)
Yhat (Emotional Symptoms Scale, age 7)	-0.01	-0.00
	(0.01)	(0.00)
Yhat (Conduct Problems Scale, age 7)	-0.01	-0.00
	(0.01)	(0.00)
Yhat (Hyperactivity Scale, age 7)	-0.00	-0.01
	(0.01)	(0.01)
Yhat (Peer Problems Scale, age 7)	-0.00	-0.00
	(0.01)	(0.00)
Yhat (Prosocial Scale, age 7)	-0.01	0.00
	(0.01)	(0.00)
Yhat (Total Difficulties Score, age 11)	-0.01	-0.01
	(0.01)	(0.01)
Yhat (Emotional Symptoms Scale, age 11)	-0.01*	-0.00
	(0.01)	(0.00)
Yhat (Conduct Problems Scale, age 11)	-0.01	-0.00
	(0.01)	(0.00)
Yhat (Hyperactivity Scale, age 11)	-0.00	-0.01
	(0.01)	(0.01)
Yhat (Peer Problems Scale, age 11)	-0.00	-0.00
	(0.01)	(0.00)
Yhat (Prosocial Scale, age 11)	-0.01	0.00
	(0.01)	(0.00)
Bandwidth	30 days	Full
Linear trend × cutoff	✓	✓
Quadratic trend × cutoff		✓

Robust standard errors in parenthesis. $^{**}p < 0.01$, $^{*}p < 0.05$. Each cell represents the point-estimate for being born on January 1st or later, from an individual regression of y-hat on this cutoff and the trends. Y-hat is created by estimating the corresponding depending variable on all covariates. 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.

provide a unambiguous judgment across outcome variables. The second order polynomial i preferred in five out of the 12 outcomes, while the first order polynomial is preferred in four cases.

Since the quadratic trend is suggested both by the graphical representation and is the most popular choice based on the AIC comparison, we apply this specification in the main regressions.

For the bandwidth selection we follow the approach taken in the existing evidence (Landersø et al., 2013) and use a 30-day bandwidth for the local regression with a linear trend interacted with cutoff. It is reasonable to assume that parents to some extent can control the childbirth by month. We also ran a regression using all children whose gestational age made it likely that they were born on January 1st. The results are very similar.

Throughout we will show results using both the local specification and the full sample specification with the parametric specified above. In the robustness analysis we show how the results vary by choice of bandwidth and parametric specification.

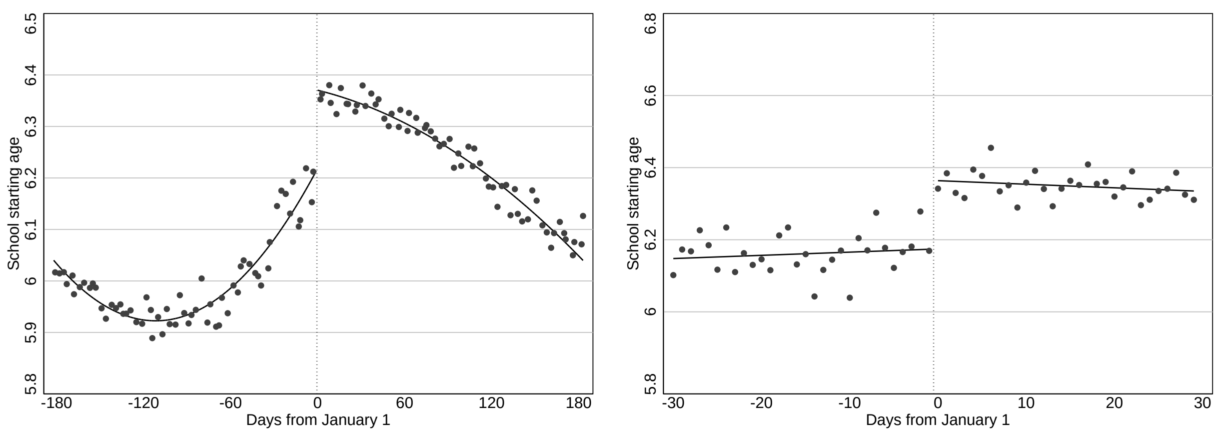
Table 7: *Model selection - Akaike Information Criteria (AIC)*

	(1)	(2)	(3)
Total Difficulties 7	-839.81	-850.40	-851.06
Emotional Symptoms 7	-217.58	-222.46	-222.53
Conduct Problems 7	-131.78	-134.61	-132.23
Hyperactivity 7	-751.81	-768.18	-766.01
Peer Problems 7	-1078.05	-1082.37	-1080.11
Pro-social Behavior 7	-674.24	-671.72	-669.85
Total Difficulties 11	-613.39	-614.93	-611.48
Emotional Symptoms 11	-188.20	-186.34	-183.78
Conduct Problems 11	-402.78	-399.14	-396.68
Hyperactivity 11	-566.61	-570.03	-570.04
Peer Problems 11	-568.71	-571.68	-568.91
Pro-social Behavior 11	-768.13	-764.28	-764.03
Linear trend × cutoff	✓	✓	✓
Quadratic trend × cutoff		✓	✓
Cubic trend × cutoff			✓

The AIC is calculated by $AIC = N * \ln(RSS/N) + 2 * k$.

5.3 Main results

The first stage relationship between date of birth and school starting age is shown in Figures 5a and 5b. For the full-year sample in Figure 5a the quadratic trends capture the variation in school starting age reasonably well. The average school starting age jumps from around 6.2 to about 6.35 at January 1st. In the local specification in Figure 5b the linear trend seems sufficient to describe the relationship. At January 1st the average school starting age jumps from 6.15 to 6.3.



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

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

Figure 5: *Date of birth and school starting age*

While Figures A.8 and A.9 in the Appendix show the "raw" reduced form relationship by plotting the mean values of the outcome variables using three day bins and fitted quadratic lines, Figures 6 and 7 show that a jump in the hyperactivity dimension can be identified even in a less parametric (i.e. no fitted polynomial) version of the reduced form relationship. The graphs show the outcome means for 15 day bins, and their corresponding 95 percent standard error confidence bands. For all outcomes, but the hyperactivity and the total difficulties score, there seems to be no jump in outcomes for children born around January 1st. But for these remaining two SDQ dimensions a discontinuity is clearly identifiable, both at age seven and age 11.

The main regression results for outcomes at age seven are shown in Table 8. Columns (1) and (2) show the results from the first stage regression. Both in the local (The upper part: Panel A.) and the parametric specification (The lower part: Panel B.), the point estimate on being born January 1st or later is positive and precisely estimated. The size of the jump is slightly smaller in the parametric specification, and adding the full set of covariates makes the jump slightly larger, but also more precisely estimated.

Columns (3) and (4) in Table 8 show coefficients from the reduced form regression of the outcome variables on the cutoff-indicator and the trends. Each cell is the point estimate on being born January 1st or later for a year running from July to June. The regressions reveal a jump in the Total Difficulties Score, which is driven by the jump on the hyperactivity scale. While adding covariates has almost no impact on the point estimates, the jumps are considerably smaller in the parametric compared to the local specifications.

The results in Table 8 stress the importance of using an instrument to obtain exogenous variation in school starting age. All OLS estimates in column (5) are significant have the reverse sign compared to the 2SLS estimates in columns (6) and (7). The OLS results show that children who enroll later have worse non-cognitive skills.

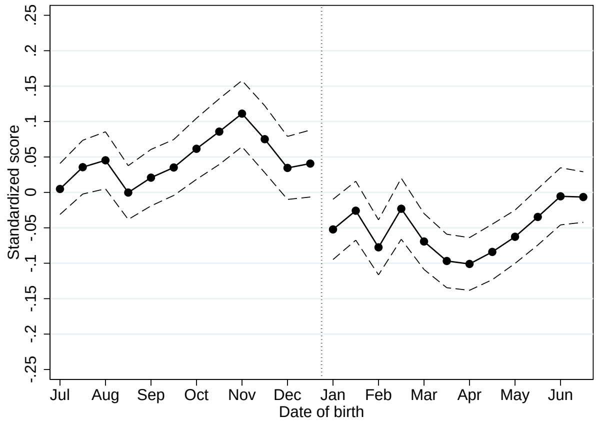
Dividing the cells in columns (3) and (4) by the cells in columns (1) and (2) gives the Wald estimates (as specified in equation (5)) of the effect of school starting age on the SDQ measures, which are presented in terms of the 2SLS coefficients in columns (6) and (7). The point estimates in the local specifications are up to 50 percent larger than the parametric specifications. The difference is driven by the fact that the reduced form jump is larger with a 30 day bandwidth than with the full sample. Enrolling in school one year later reduces hyperactivity by between 0.5 and 0.8 of a standard deviation, while the other dimensions remain unaffected.

Table 9 shows estimation results for outcomes at age 11. The effects are very similar to the age seven effects. Children who are older at enrollment have fewer difficulties at age 11, and this seems to be driven mainly by less hyperactivity problems. At age 11 there are also signs of fewer Peer Problems.

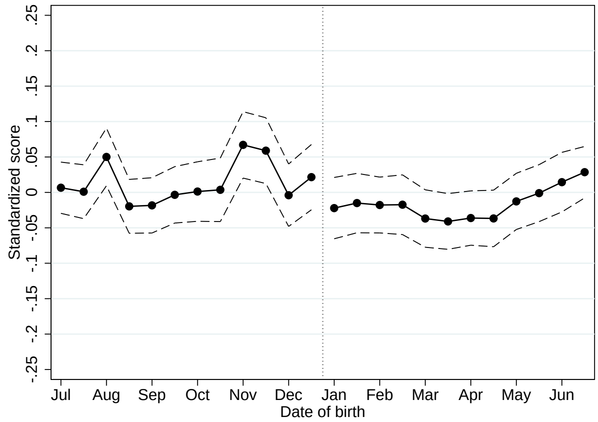
Recall that the effect is only identified for the subgroup of the population that complies to the school starting age rule. Table 10 provides a comparison of the first-stage coefficients for subgroups of the population. Compliers are mostly children of low educated parents, girls and children with no older siblings.

5.4 Heterogeneity

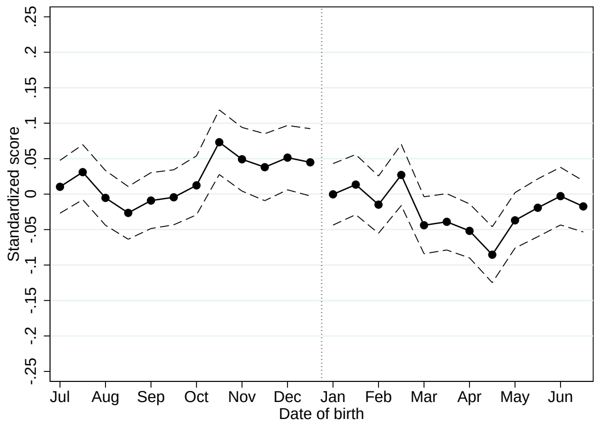
Tables 11 and 12 show the results divided by gender. As we already saw in Table 10 the first stage is much stronger for girls than for boys. The results in the table reveal that the girls also have slightly larger reduced form point estimates, but the difference in the first stage is also large, so that the 2SLS estimates are much less precise for boys.



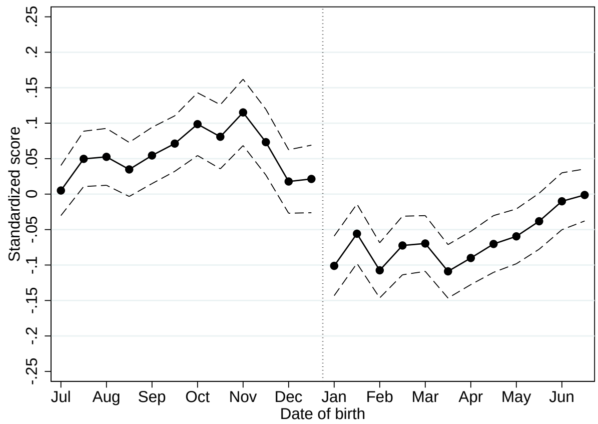
(a) *Total difficulties score*



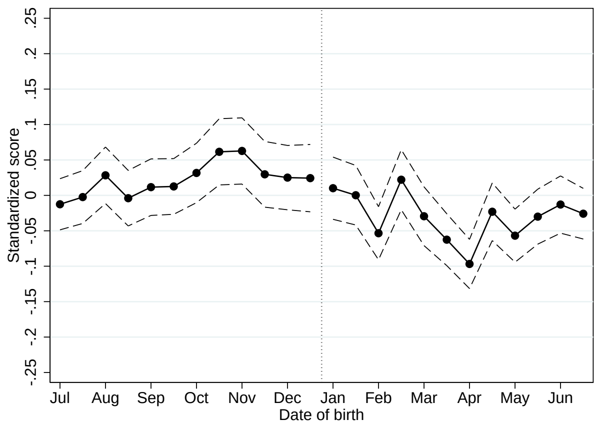
(b) *Conduct problems*



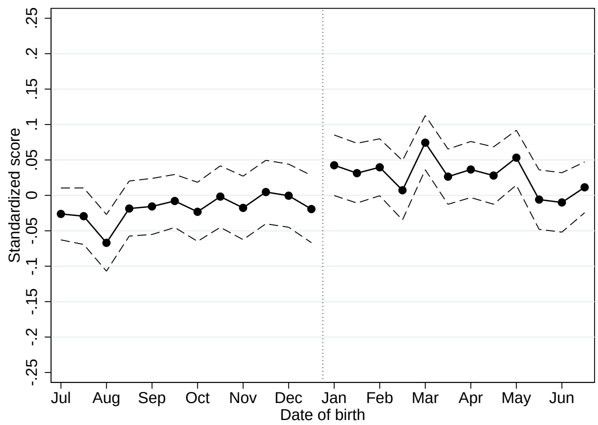
(c) *Emotional symptoms*



(d) *Hyperactivity*

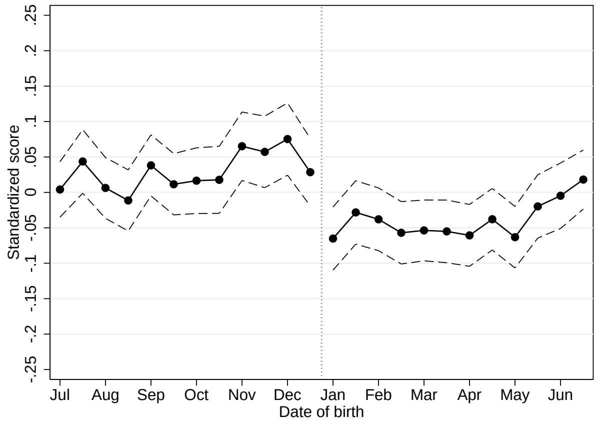


(e) *Peer problems*

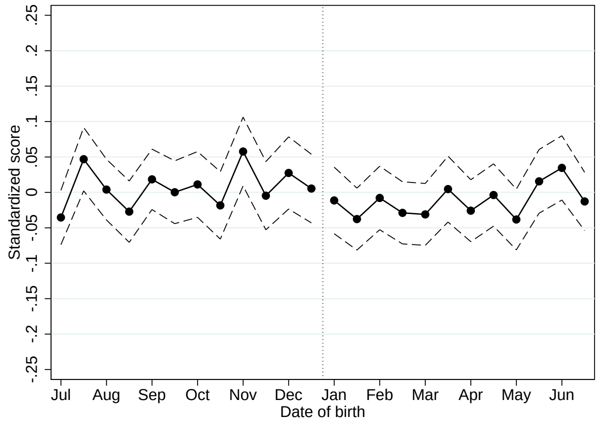


(f) *Pro-social*

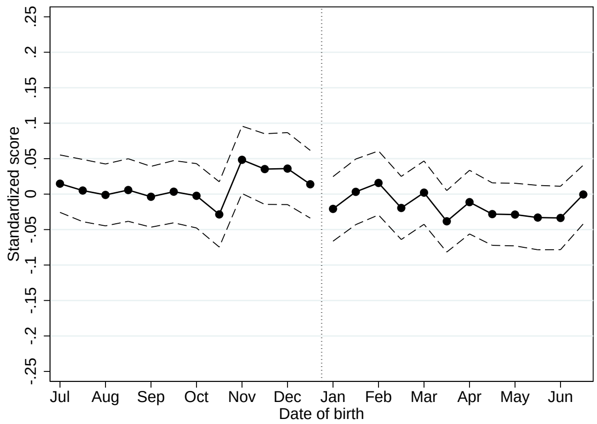
Figure 6: *SDQ outcomes at age 7 and date of birth. All outcome variables are standardized. Full year, 15 day averages.*



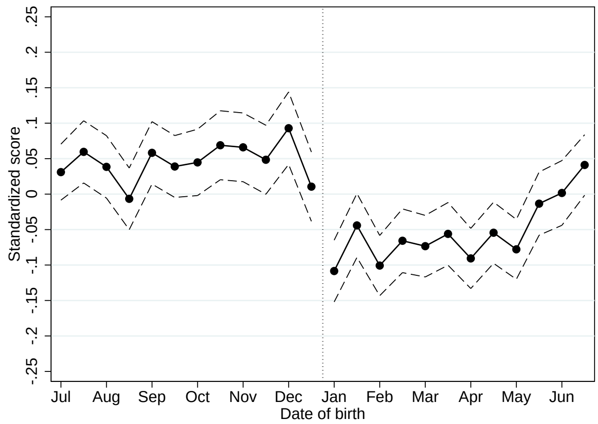
(a) *Total difficulties score*



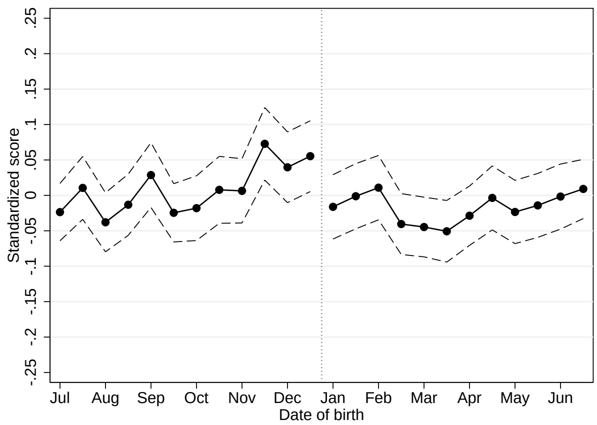
(b) *Conduct problems*



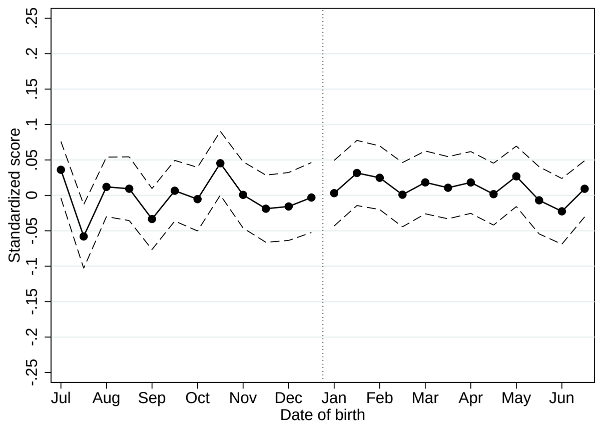
(c) *Emotional symptoms*



(d) *Hyperactivity*



(e) *Peer problems*



(f) *Pro-social*

Figure 7: *SDQ outcomes at age 11 and date of birth. All outcome variables are standardized. Full year, 15 day averages.*

Table 8: *The first stage, reduced form, and 2SLS results for SDQ outcomes age 7*

	First stage		Reduced form		OLS	2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Local specification: 30 day bandwidth & linear trends interacted with January 1. cutoff							
Birthday ≥ Jan. 1	0.16***	0.17***					
	(0.03)	(0.03)					
Total Difficulties 7			-0.13***	-0.12***	0.15***	-0.81**	-0.71**
			(0.05)	(0.05)	(0.02)	(0.33)	(0.30)
Emotional Symptoms 7			-0.08*	-0.07	0.10***	-0.50	-0.41
			(0.05)	(0.05)	(0.02)	(0.31)	(0.29)
Conduct Problems 7			-0.05	-0.05	0.08***	-0.34	-0.28
			(0.05)	(0.05)	(0.02)	(0.30)	(0.28)
Hyperactivity 7			-0.15***	-0.14***	0.14***	-0.93***	-0.87***
			(0.05)	(0.05)	(0.02)	(0.34)	(0.31)
Peer Problems 7			-0.04	-0.03	0.06***	-0.25	-0.20
			(0.05)	(0.05)	(0.02)	(0.30)	(0.29)
Pro-social Behavior 7			0.08*	0.09**	-0.11***	0.52*	0.54*
			(0.05)	(0.05)	(0.02)	(0.31)	(0.29)
Observations	7,717	7,717	7,717	7,717	7,717	7,717	7,717
B. Parametric specification: Full sample & quadratic trends interacted with January 1. cutoff							
Birthday ≥ Jan. 1	0.14***	0.14***					
	(0.02)	(0.01)					
Total Difficulties 7			-0.08***	-0.07***	0.21***	-0.58***	-0.49**
			(0.03)	(0.03)	(0.01)	(0.21)	(0.19)
Emotional Symptoms 7			-0.05*	-0.04	0.11***	-0.35*	-0.29
			(0.03)	(0.03)	(0.01)	(0.20)	(0.19)
Conduct Problems 7			-0.04	-0.04	0.13***	-0.31	-0.27
			(0.03)	(0.03)	(0.01)	(0.19)	(0.19)
Hyperactivity 7			-0.09***	-0.08***	0.20***	-0.66***	-0.59***
			(0.03)	(0.03)	(0.01)	(0.21)	(0.20)
Peer Problems 7			-0.02	-0.01	0.13***	-0.14	-0.08
			(0.03)	(0.03)	(0.01)	(0.19)	(0.19)
Pro-social Behavior 7			0.04	0.04	-0.12***	0.26	0.26
			(0.03)	(0.03)	(0.01)	(0.19)	(0.19)
Observations	55,026	55,026	55,026	55,026	55,026	55,026	55,026
Covariates		✓		✓			✓

Robust standard errors in parenthesis. $^{***}p < 0.01$ $^{**}p < 0.05$, $^{*}p < 0.1$. Each cell shows the estimate from a single regression. In columns (1) and (2) the dependent variable is school starting age (in years) which is regressed on an indicator for being born after January 1st, trends, and trends interacted with the January 1st cutoff. In columns (3) and (4) the SDQ measure is regressed on the same specification as in (1) and (2). Column (5) shows the results from a simple OLS regression of the dependent variable on school starting age and the time trends. Columns (6) and (7) show the 2SLS results from estimating the SDQ measure on the predicted school starting age, the time trends, and the time trends interacted with the cutoff. 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.

Table 9: *The first stage, reduced form, and 2SLS results for SDQ outcomes age 11*

	First stage		Reduced form		OLS	2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Local specification: 30 day bandwidth & linear trends interacted with January 1. cutoff							
Birthday ≥ Jan. 1	0.19***	0.20***					
	(0.03)	(0.03)					
Total Difficulties 11			-0.10**	-0.08*	0.15***	-0.51*	-0.42*
			(0.05)	(0.05)	(0.02)	(0.26)	(0.25)
Emotional Symptoms 11			-0.04	-0.03	0.07***	-0.22	-0.13
			(0.05)	(0.05)	(0.03)	(0.25)	(0.25)
Conduct Problems 11			-0.01	0.00	0.12***	-0.08	0.02
			(0.05)	(0.05)	(0.02)	(0.25)	(0.25)
Hyperactivity 11			-0.10**	-0.10**	0.15***	-0.54**	-0.50**
			(0.05)	(0.05)	(0.02)	(0.26)	(0.25)
Peer Problems 11			-0.10**	-0.10*	0.07***	-0.53**	-0.49*
			(0.05)	(0.05)	(0.02)	(0.27)	(0.26)
Pro-social Behavior 11			0.04	0.05	-0.12***	0.23	0.26
			(0.05)	(0.05)	(0.02)	(0.26)	(0.25)
Observations	6,812	6,812	6,812	6,812	6,812	6,812	6,812
B. Parametric specification: Full sample & quadratic trends interacted with January 1. cutoff							
Birthday ≥ Jan. 1	0.16***	0.16***					
	(0.02)	(0.02)					
Total Difficulties 11			-0.10***	-0.09***	0.20***	-0.64***	-0.55***
			(0.03)	(0.03)	(0.01)	(0.19)	(0.18)
Emotional Symptoms 11			-0.04	-0.03	0.10***	-0.26	-0.19
			(0.03)	(0.03)	(0.01)	(0.18)	(0.18)
Conduct Problems 11			-0.04	-0.02	0.12***	-0.22	-0.15
			(0.03)	(0.03)	(0.01)	(0.18)	(0.18)
Hyperactivity 11			-0.12***	-0.11***	0.20***	-0.76***	-0.67***
			(0.03)	(0.03)	(0.01)	(0.20)	(0.18)
Peer Problems 11			-0.08***	-0.07**	0.13***	-0.47**	-0.43**
			(0.03)	(0.03)	(0.01)	(0.19)	(0.19)
Pro-social Behavior 11			0.03	0.03	-0.09***	0.19	0.18
			(0.03)	(0.03)	(0.01)	(0.18)	(0.18)
Observations	45,630	45,630	45,630	45,630	45,630	45,630	45,630
Covariates		✓		✓			✓

Robust standard errors in parenthesis. $^{***}p < 0.01$ $^{**}p < 0.05$, $^{*}p < 0.1$. Each cell shows the estimate from a single regression. In columns (1) and (2) the dependent variable is school starting age (in years) which is regressed on an indicator for being born after January 1st, trends, and trends interacted with the January 1st cutoff. In columns (3) and (4) the SDQ measure is regressed on the same specification as in (1) and (2). Column (5) shows the results from a simple OLS regression of the dependent variable on school starting age and the time trends. Columns (6) and (7) show the 2SLS results from estimating the SDQ measure on the predicted school starting age, the time trends, and the time trends interacted with the cutoff. 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.

Table 10: *Characterizing compliers, comparison of the first stage across subgroups of the sample.*

	(1)	(2)	(3)	(4)	(5)	(6)
	Girls	Boys	No siblings	Siblings	Low educ.	Highly educ.
Coefficient	0.21	0.16	0.22	0.15	0.21	0.17
t-value	(6.05)	(5.34)	(7.10)	(4.22)	(6.51)	(5.16)
Relative to overall first stage	1.09	0.85	1.17	0.79	1.13	0.90
Observations	4,696	4,744	5,207	4,233	4,524	4,916
Bandwidth	30 d	30 d	30 d	30 d	30 d	30 d
Linear trend × cut	✓	✓	✓	✓	✓	✓

The row "Coefficient" gives the point-estimate on being born on January 1st or later, in a regression of school starting age on this indicator, the trend, the trend interacted with the cutoff, and the covariates. The t-value is the t-statistic on this estimate. The third rows shows the ratio of the subgroup point-estimate to the full sample point-estimate of the first stage. High/low education is measured by more or less that 15 years of education.

Table 11: *The first stage, reduced form, and 2SLS results by gender, age 7*

	First stage		Reduced form		2SLS	
	Boys	Girls	Boys	Girls	Boys	Girls
	(1)	(2)	(3)	(4)	(5)	(6)
A. Local specification: 30 day bandwidth & linear trends interacted with January 1. cutoff						
Birthday $\geq$ Jan. 1	0.14***	0.19***				
	(0.03)	(0.04)				
Total Difficulties 7			-0.07	-0.18***	-0.48	-0.93**
			(0.07)	(0.06)	(0.52)	(0.37)
Emotional Symptoms 7			0.01	-0.14**	0.05	-0.75*
			(0.07)	(0.07)	(0.48)	(0.39)
Conduct Problems 7			-0.05	-0.05	-0.35	-0.28
			(0.07)	(0.06)	(0.50)	(0.33)
Hyperactivity 7			-0.13*	-0.17***	-0.92	-0.87**
			(0.07)	(0.06)	(0.56)	(0.36)
Peer Problems 7			0.02	-0.10	0.17	-0.50
			(0.07)	(0.06)	(0.52)	(0.33)
Pro-social Behavior 7			0.04	0.14**	0.31	0.74**
			(0.07)	(0.06)	(0.51)	(0.35)
Observations	3,912	3,805	3,912	3,805	3,912	3,805
B. Parametric specification: Full sample & quadratic trends interacted with January 1. cutoff						
Birthday $\geq$ Jan. 1	0.06***	0.23***				
	(0.02)	(0.02)				
Total Difficulties 7			-0.03	-0.12***	-0.44	-0.51***
			(0.04)	(0.03)	(0.68)	(0.16)
Emotional Symptoms 7			-0.00	-0.08**	-0.07	-0.35**
			(0.04)	(0.04)	(0.62)	(0.17)
Conduct Problems 7			-0.03	-0.05	-0.50	-0.21
			(0.04)	(0.04)	(0.67)	(0.16)
Hyperactivity 7			-0.05	-0.12***	-0.82	-0.53***
			(0.04)	(0.03)	(0.72)	(0.16)
Peer Problems 7			0.03	-0.05	0.43	-0.22
			(0.04)	(0.04)	(0.68)	(0.16)
Pro-social Behavior 7			0.02	0.05	0.37	0.23
			(0.04)	(0.03)	(0.68)	(0.15)
Observations	28,106	26,920	28,106	26,920	28,106	26,920
Covariates	✓	✓	✓	✓	✓	✓

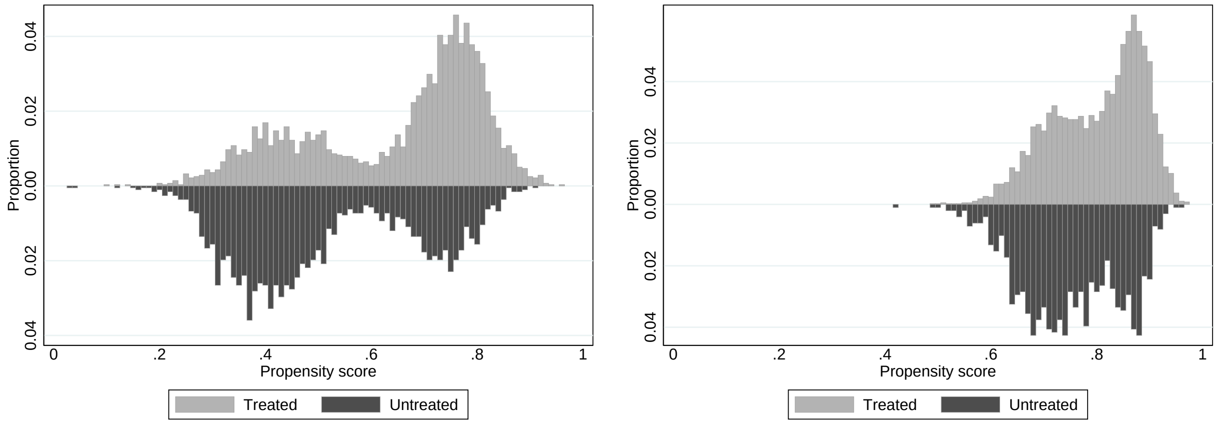
Robust standard errors in parenthesis. $^{***}p < 0.01$ $^{**}p < 0.05$, $^{*}p < 0.1$.Each cell shows the estimate from a single regression. In columns (1) and (2) the dependent variable is school starting age (in years) which is regressed on an indicator for being born after January 1st, trends, and trends interacted with the January 1st cutoff. In columns (3) and (4) the SDQ measure is regressed on the same specification as in (1) and (2). Columns (5) and (6) show the 2SLS results from estimating the SDQ measure on the predicted school starting age, the time trends, and the time trends interacted with the cutoff. 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.

Table 12: *The first stage, reduced form, and 2SLS results by gender, age 11*

	First stage		Reduced form		2SLS	
	Boys	Girls	Boys	Girls	Boys	Girls
	(1)	(2)	(3)	(4)	(5)	(6)
A. Local specification: 30 day bandwidth & linear trends interacted with January 1. cutoff						
Birthday ≥ Jan. 1	0.18***	0.20***				
	(0.04)	(0.04)				
Total Difficulties 11			-0.02	-0.12**	-0.13	-0.61*
			(0.07)	(0.06)	(0.40)	(0.33)
Emotional Symptoms 11			0.06	-0.09	0.31	-0.44
			(0.07)	(0.07)	(0.37)	(0.35)
Conduct Problems 11			0.06	-0.04	0.31	-0.21
			(0.07)	(0.06)	(0.41)	(0.32)
Hyperactivity 11			-0.07	-0.11*	-0.42	-0.53*
			(0.07)	(0.06)	(0.42)	(0.30)
Peer Problems 11			-0.08	-0.09	-0.46	-0.46
			(0.08)	(0.06)	(0.43)	(0.33)
Pro-social Behavior 11			0.05	0.05	0.29	0.23
			(0.08)	(0.06)	(0.43)	(0.31)
Observations	3,363	3,449	3,363	3,449	3,363	3,449
B. Parametric specification: Full sample & quadratic trends interacted with January 1. cutoff						
Birthday ≥ Jan. 1	0.08***	0.25***				
	(0.02)	(0.02)				
Total Difficulties 11			-0.05	-0.13***	-0.61	-0.53***
			(0.04)	(0.04)	(0.60)	(0.16)
Emotional Symptoms 11			0.03	-0.09**	0.42	-0.37**
			(0.04)	(0.04)	(0.53)	(0.17)
Conduct Problems 11			-0.03	-0.02	-0.34	-0.10
			(0.04)	(0.04)	(0.58)	(0.15)
Hyperactivity 11			-0.06	-0.15***	-0.84	-0.62***
			(0.04)	(0.04)	(0.63)	(0.15)
Peer Problems 11			-0.08*	-0.06	-1.03	-0.25
			(0.04)	(0.04)	(0.66)	(0.16)
Pro-social Behavior 11			0.03	0.03	0.43	0.11
			(0.04)	(0.04)	(0.60)	(0.15)
Observations	22,685	22,945	22,685	22,945	22,685	22,945
Covariates	✓	✓	✓	✓	✓	✓

Robust standard errors in parenthesis. $^{***}p < 0.01$ $^{**}p < 0.05$, $^{*}p < 0.1$. Each cell shows the estimate from a single regression. In columns (1) and (2) the dependent variable is school starting age (in years) which is regressed on an indicator for being born after January 1st, trends, and trends interacted with the January 1st cutoff. In columns (3) and (4) the SDQ measure is regressed on the same specification as in (1) and (2). Columns (5) and (6) show the 2SLS results from estimating the SDQ measure on the predicted school starting age, the time trends, and the time trends interacted with the cutoff. 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.

While subgroup regressions provide a means to compare differences in treatment effects across observables, marginal treatment effects allow us to compare treatment effects along an "unobserved" dimension. To compute the marginal treatment effects we consider the local bandwidth of 30 days and consider a binary treatment variable, SSO, which takes the value of one if the individual is older than 6.5 years at school enrollment. Using this specification we calculate the marginal treatment effects by means of the parametric specification outlined in Heckman et al. (2006) and used by Landersø et al. (2013).



(a) *Girls*

(b) *Boys*

Figure 8: *Common support. Treatment: School Starting Age > 6.5. 30 days bandwidth. We estimated a probit model where the dependent variables takes the value of one, if the child is older than 6.5 years at enrollment. Covariates included are birth weight, 5 minute APGAR score, parental education, parents' age, parental income, parental employment, age at test (monthly indicators), linear time trends, and birth year fixed effects.*

We first estimate a probit model for being treated, and predict each individual's propensity score for being old at school enrollment. This propensity score is then included in a specification for the outcome equation. Intuitively for every observed level of probability for treatment (the propensity score) we compare individuals who were treated to individuals who were not treated. The effect can therefore only be identified for propensity scores where we observe both treated and untreated children. Figure 8 shows that for girls there is a common support over the interval 0.2 to 0.9. For boys there is essentially no one with a propensity score below 0.6. Very few boys start early. This is in line with the previous section where we found that compliers are primarily girls.

Children with a very high level of observed probability who were not treated must have had a large error term to not make them treated. This is what Landersø et al. (2013) interpret as latent utility, u. Figures 9 and 10 show the marginal treatment effects for all outcomes for girls plotted against u. Although the confidence bands are relatively wide, an upward sloping trend is recognizable for most outcomes. The marginal treatment effect is only significant for the lower half of the "latent ability" distribution: Primarily girls with a low level of latent utility benefit from enrolling in school later, measured by the Total Difficulties score and the hyperactivity scale. This finding is somewhat in contrast to the conclusion by Landersø et al. (2013), who find that the marginal treatment effects of school starting age on crime are homogeneous for girls. It makes intuitive sense, that the estimated treatment effect is largest for girls with a low level of latent ability, while for girls with a high level of latent ability starting school later has no benefits.

For boys the marginal treatment effects is identified for a much narrower bandwidth covering the propensity scores from 0.6 to 0.9. While for girls most treatment effects were decreasing in latent ability (because lower values means larger effects), Figures 11 and 12 reveal that for boys the reverse is the case. The confidence bands are also quite wide for boys, but especially for the hyperactivity scale at age seven we find signs of a downward sloping relationship. Only boys with a very high level of latent ability benefit from enrolling late in school. This is in line with the finding by Landersø et al. (2013) who conclude that older school starting age reduces crime for the most able boys.

5.5 Robustness

As the outcome variables are parent reported answers to the Strength and Difficulties Questionnaire, the effect may stem from the fact that parents compare their children to the class mates. It could therefore be a pure measurement effect. To assess this we consider the subsample of children who have older siblings. The intuition is that for children who have older siblings the parent should have another reference category than the class mates. If we find no effect for this subgroup, the main effect could be driven by pure measurement. This is not the case as shown by Tables 13 and 14, in contrast the point estimates of the reduced form regression and the 2SLS regressions are all larger for children with older siblings. Including the average age of the cohort (excluding the individual) has a small impact in the local specification on the age 11 outcomes as Tables A.5 and A.6 in the Appendix show, but in all other cases the conclusion remains unaffected.

In Figures 13 and 14 we assess the sensitivity to the bandwidth selection, and the parametric specification. Focusing on the hyperactivity scale we find that using a first, second or third order polynomial does not affect the point estimate notably. Also for almost all bandwidths the local specification coefficients are in line with the parametric first and second order polynomial specification.

Table 13: *The first stage, reduced form, and 2SLS results by sibling status, age 7*

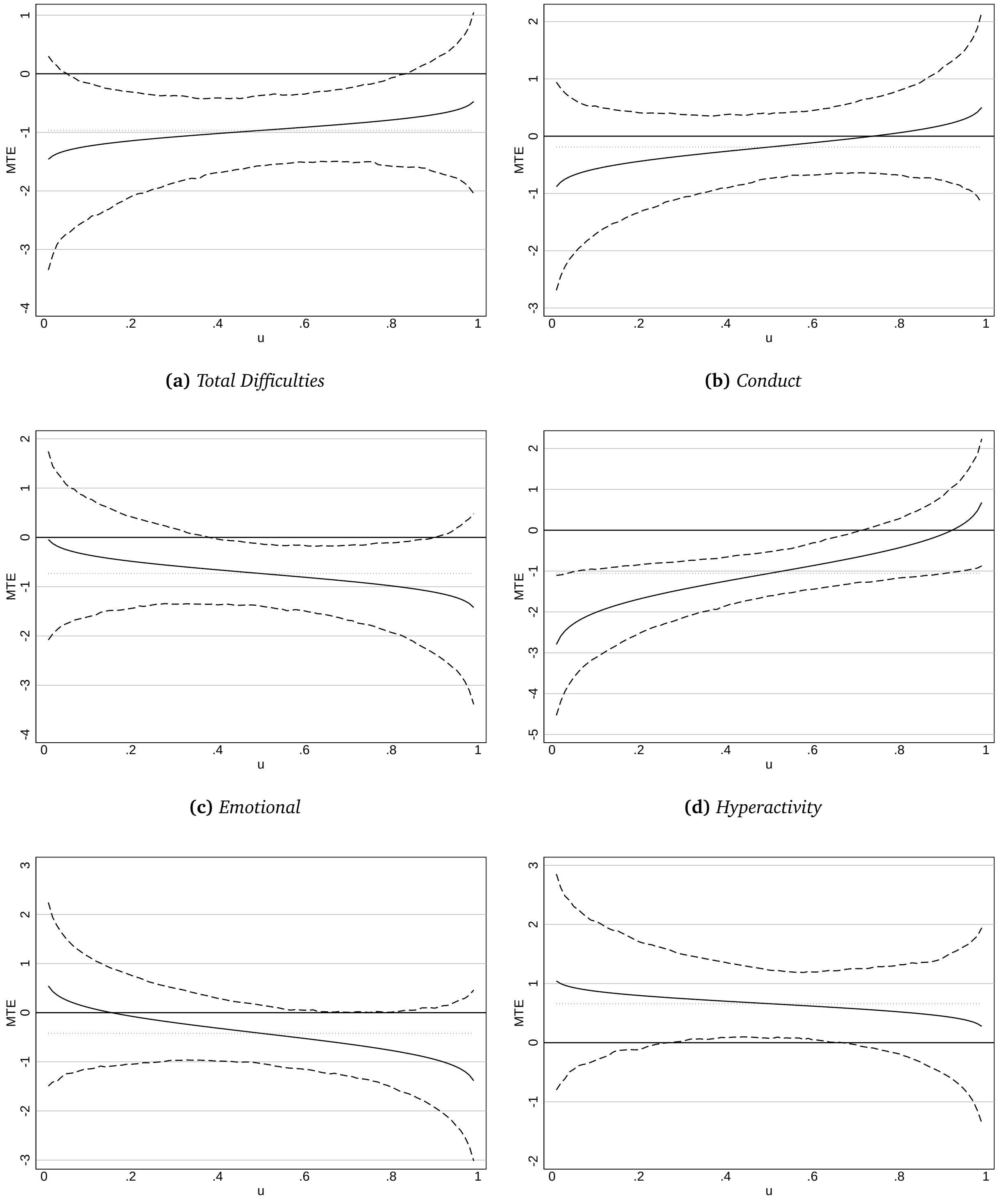
	First stage		Reduced form		2SLS	
Older Siblings	No	Yes	No	Yes	No	Yes
	(1)	(2)	(3)	(4)	(5)	(6)
A. Local specification: 30 day bandwidth & linear trends interacted with January 1. cutoff						
Birthday ≥ Jan. 1	0.19***	0.14***				
	(0.03)	(0.04)				
Total Difficulties 7			-0.07	-0.17**	-0.38	-1.24**
			(0.06)	(0.07)	(0.34)	(0.63)
Emotional Symptoms 7			0.00	-0.14**	0.01	-1.05*
			(0.07)	(0.07)	(0.35)	(0.58)
Conduct Problems 7			-0.01	-0.09	-0.06	-0.67
			(0.06)	(0.07)	(0.32)	(0.55)
Hyperactivity 7			-0.13**	-0.15**	-0.67*	-1.13*
			(0.06)	(0.07)	(0.35)	(0.61)
Peer Problems 7			-0.03	-0.05	-0.14	-0.36
			(0.06)	(0.07)	(0.33)	(0.54)
Pro-social Behavior 7			0.07	0.11	0.37	0.85
			(0.06)	(0.07)	(0.33)	(0.58)
Observations	4,293	3,424	4,293	3,424	4,293	3,424
B. Parametric specification: Full sample & Quadratic trends interacted with January 1. cutoff						
Birthday ≥ Jan. 1	0.19***	0.14***				
	(0.03)	(0.04)				
Total Difficulties 7			-0.07	-0.17**	-0.38	-1.24**
			(0.06)	(0.07)	(0.34)	(0.63)
Emotional Symptoms 7			0.00	-0.14**	0.01	-1.05*
			(0.07)	(0.07)	(0.35)	(0.58)
Conduct Problems 7			-0.01	-0.09	-0.06	-0.67
			(0.06)	(0.07)	(0.32)	(0.55)
Hyperactivity 7			-0.13**	-0.15**	-0.67*	-1.13*
			(0.06)	(0.07)	(0.35)	(0.61)
Peer Problems 7			-0.03	-0.05	-0.14	-0.36
			(0.06)	(0.07)	(0.33)	(0.54)
Pro-social Behavior 7			0.07	0.11	0.37	0.85
			(0.06)	(0.07)	(0.33)	(0.58)
Observations	4,293	3,424	4,293	3,424	4,293	3,424
Covariates	✓	✓	✓	✓	✓	✓

Robust standard errors in parenthesis. $^{***}p < 0.01$ $^{**}p < 0.05$, $^{*}p < 0.1$. Each cell shows the estimate from a single regression. In columns (1) and (2) the dependent variable is school starting age (in years) which is regressed on an indicator for being born after January 1st, trends, and trends interacted with the January 1st cutoff. In columns (3) and (4) the SDQ measure is regressed on the same specification as in (1) and (2). Column (5) shows the results from a simple OLS regression of the dependent on variable on school starting age and the time trends. Columns (5) and (6) show the 2SLS results from estimating the SDQ measure on the predicted school starting age, the time trends, and the time trends interacted with the cutoff. 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.

Table 14: *The first stage, reduced form, and 2SLS results by sibling status, age 11*

	First stage		Reduced form		2SLS	
Older Siblings	No	Yes	No	Yes	No	Yes
	(1)	(2)	(3)	(4)	(5)	(6)
A. Local specification: 30 day bandwidth & linear trends interacted with January 1. cutoff						
Birthday ≥ Jan. 1	0.23***	0.16***				
	(0.03)	(0.04)				
Total Difficulties 11			0.01	-0.20***	0.06	-1.23**
			(0.06)	(0.07)	(0.28)	(0.55)
Emotional Symptoms 11			0.03	-0.09	0.13	-0.56
			(0.07)	(0.07)	(0.29)	(0.46)
Conduct Problems 11			0.08	-0.10	0.37	-0.63
			(0.07)	(0.07)	(0.29)	(0.49)
Hyperactivity 11			-0.03	-0.18**	-0.14	-1.11**
			(0.06)	(0.07)	(0.28)	(0.53)
Peer Problems 11			-0.02	-0.19***	-0.08	-1.16**
			(0.07)	(0.07)	(0.30)	(0.55)
Pro-social Behavior 11			-0.03	0.13*	-0.11	0.82
			(0.07)	(0.07)	(0.29)	(0.51)
Observations	3,739	3,073	3,739	3,073	3,739	3,073
B. Parametric specification: Full sample & Quadratic trends interacted with January 1. cutoff						
Birthday ≥ Jan. 1	0.23***	0.16***				
	(0.03)	(0.04)				
Total Difficulties 11			0.01	-0.20***	0.06	-1.23**
			(0.06)	(0.07)	(0.28)	(0.55)
Emotional Symptoms 11			0.03	-0.09	0.13	-0.56
			(0.07)	(0.07)	(0.29)	(0.46)
Conduct Problems 11			0.08	-0.10	0.37	-0.63
			(0.07)	(0.07)	(0.29)	(0.49)
Hyperactivity 11			-0.03	-0.18**	-0.14	-1.11**
			(0.06)	(0.07)	(0.28)	(0.53)
Peer Problems 11			-0.02	-0.19***	-0.08	-1.16**
			(0.07)	(0.07)	(0.30)	(0.55)
Pro-social Behavior 11			-0.03	0.13*	-0.11	0.82
			(0.07)	(0.07)	(0.29)	(0.51)
Observations	3,739	3,073	3,739	3,073	3,739	3,073
Covariates	✓	✓	✓	✓	✓	✓

Robust standard errors in parenthesis. $^{***}p < 0.01$ $^{**}p < 0.05$, $^{*}p < 0.1$. Each cell shows the estimate from a single regression. In columns (1) and (2) the dependent variable is school starting age (in years) which is regressed on an indicator for being born after January 1st, trends, and trends interacted with the January 1st cutoff. In columns (3) and (4) the SDQ measure is regressed on the same specification as in (1) and (2). Column (5) shows the results from a simple OLS regression of the dependent on variable on school starting age and the time trends. Columns (5) and (6) show the 2SLS results from estimating the SDQ measure on the predicted school starting age, the time trends, and the time trends interacted with the cutoff. 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.

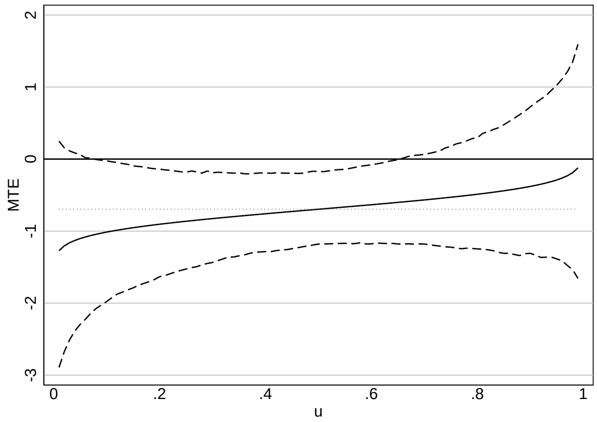


(a) *Total Difficulties* **(b)** *Conduct*

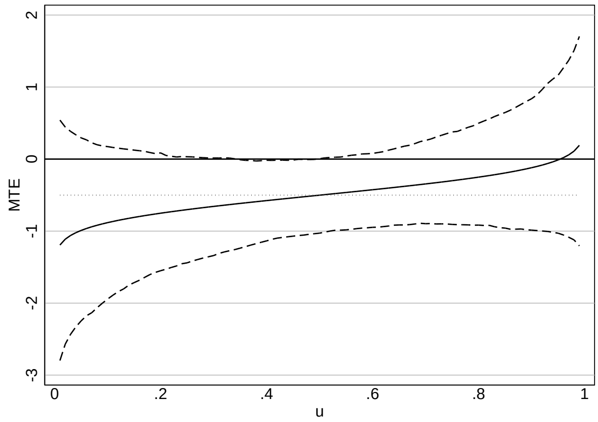
(c) *Emotional* **(d)** *Hyperactivity*

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

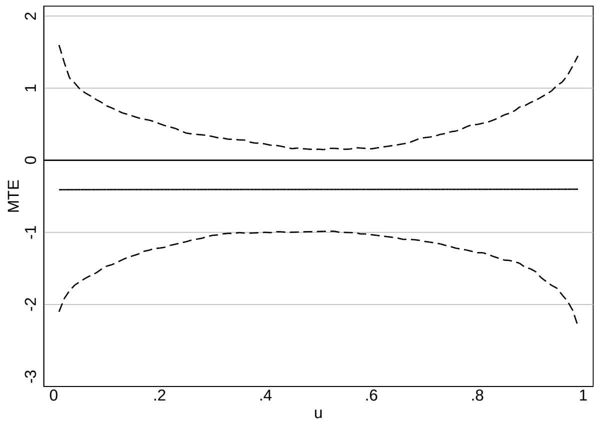
Figure 9: *Marginal treatment effects for girls, age 7. Treatment: School Starting Age > 6.5. Sample: 30 days bandwidth. We use a parametric specification in which we first estimated a probit for the propensity of being old at school enrollment, and then computed the marginal treatment effects according to Heckman et al. (2006). Bootstrapped standard errors are computed using 1,000 replications. The dotted lines indicate the average treatment effect.*



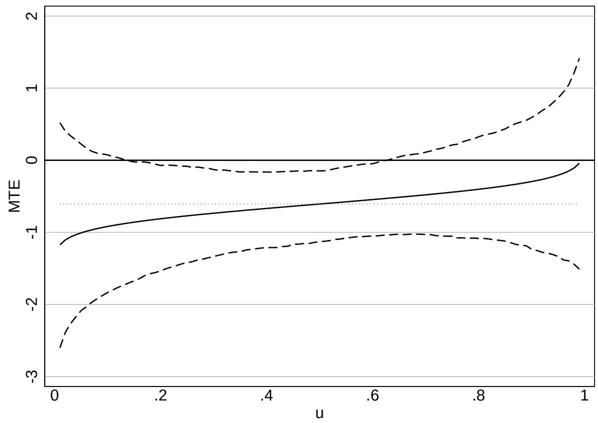
(a) *Total Difficulties*



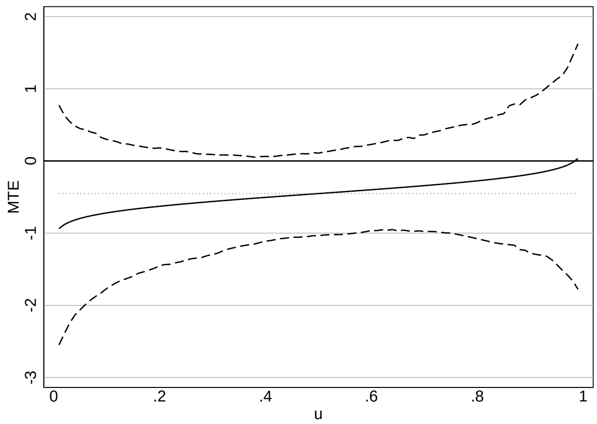
(b) *Conduct*



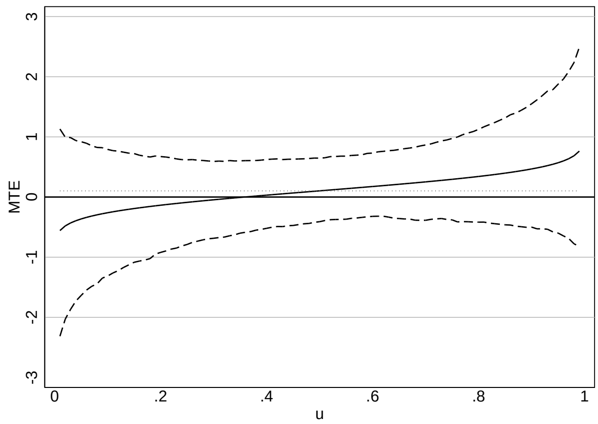
(c) *Emotional*



(d) *Hyperactivity*

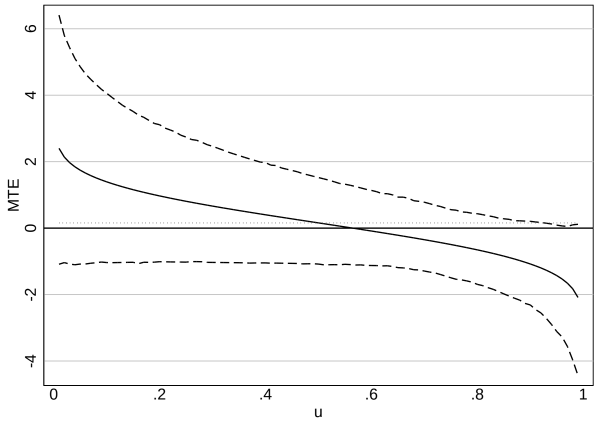


(e) *Peer problems*

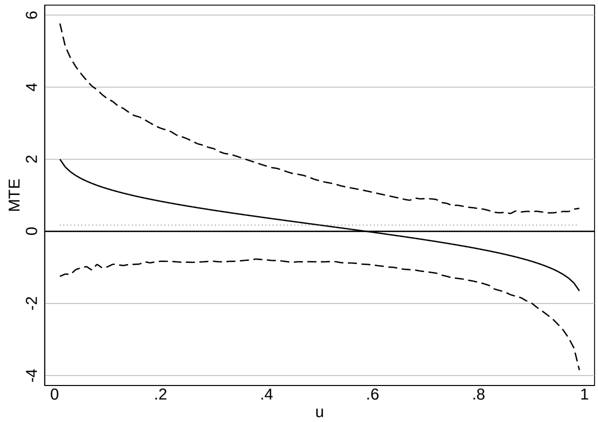


(f) *Pro-social*

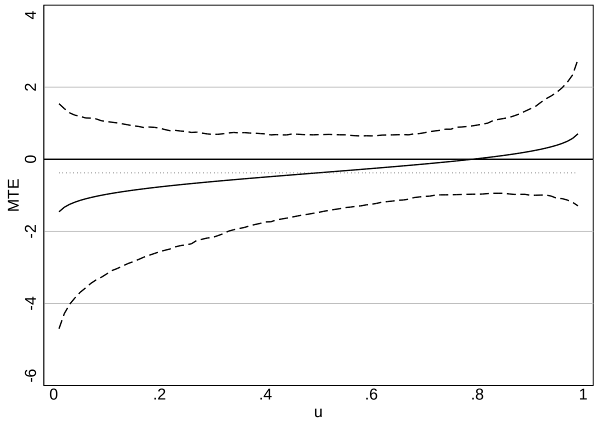
Figure 10: *Marginal treatment effects for girls, age 11. Treatment: School Starting Age > 6.5. Sample: 30 days bandwidth. We use a parametric specification in which we first estimated a probit for the propensity of being old at school enrollment, and then computed the marginal treatment effects according to Heckman et al. (2006). Bootstrapped standard errors are computed using 1,000 replications. The dotted lines indicate the average treatment effect.*



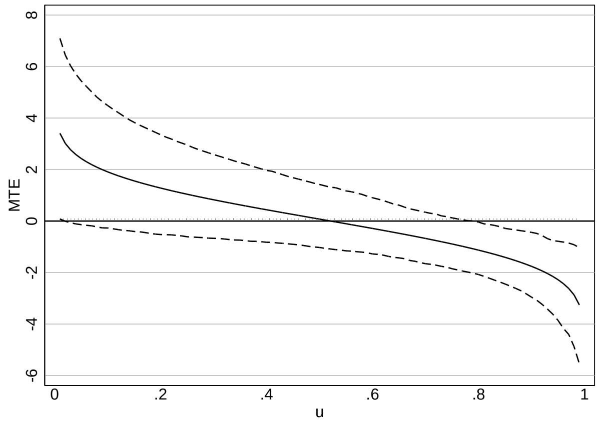
(a) *Total Difficulties*



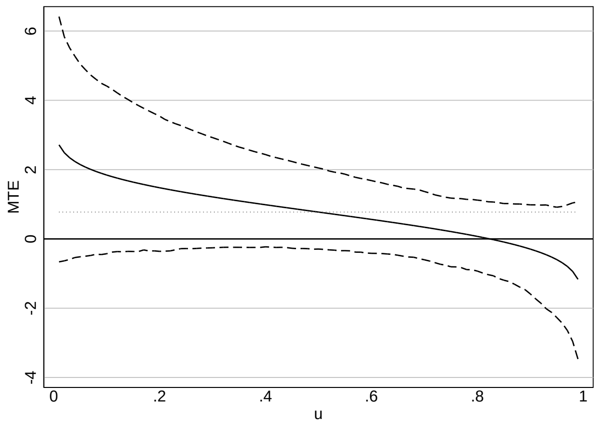
(b) *Conduct*



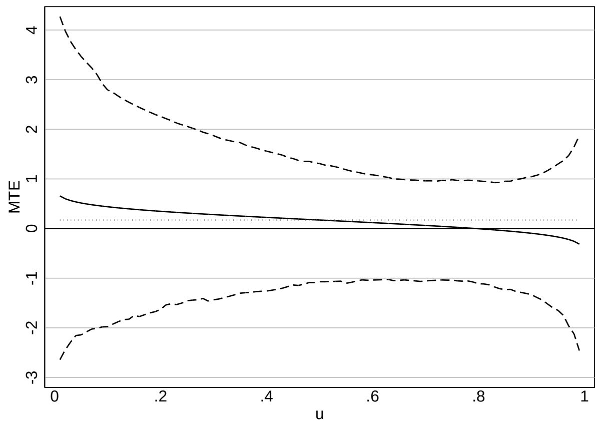
(c) *Emotional*



(d) *Hyperactivity*

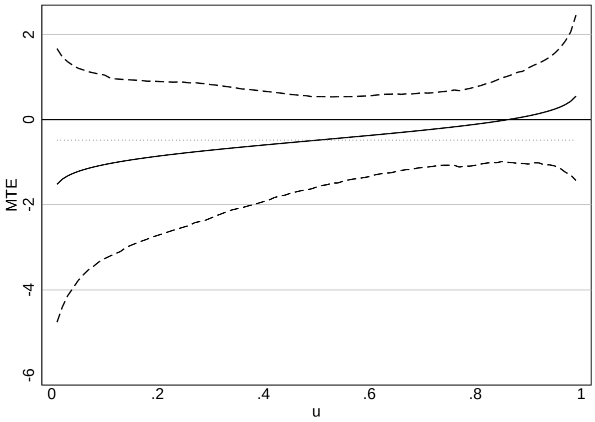


(e) *Peer problems*

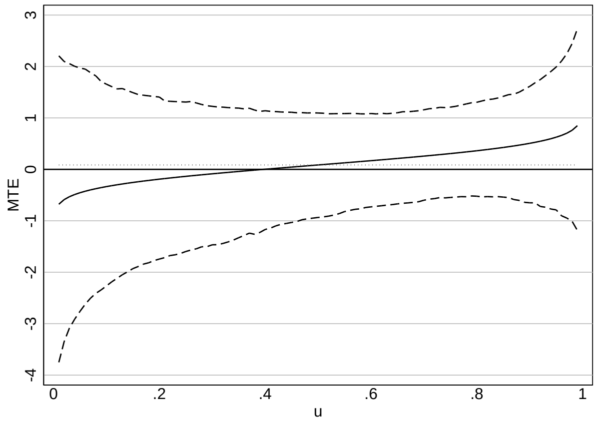


(f) *Pro-social*

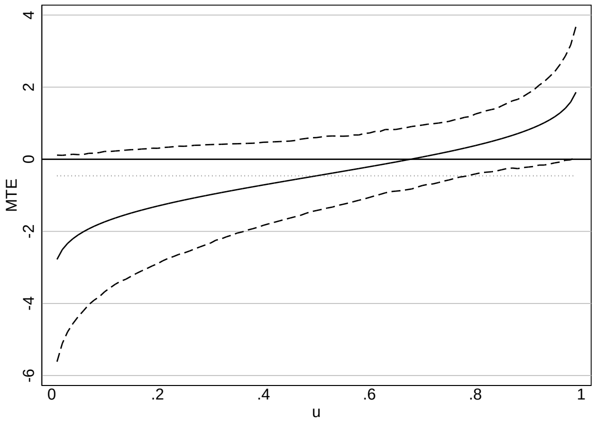
Figure 11: *Marginal treatment effects for boys, age 7. Treatment: School Starting Age > 6.5. Sample: 30 days bandwidth. We use a parametric specification in which we first estimated a probit for the propensity of being old at school enrollment, and then computed the marginal treatment effects according to Heckman et al. (2006). Bootstrapped standard errors are computed using 1,000 replications. The dotted lines indicate the average treatment effect.*



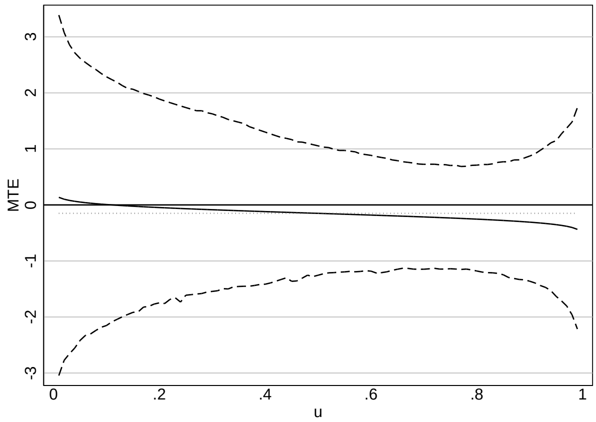
(a) *Total Difficulties*



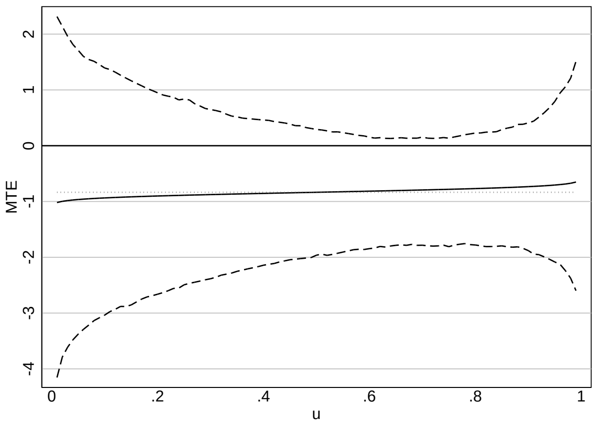
(b) *Conduct*



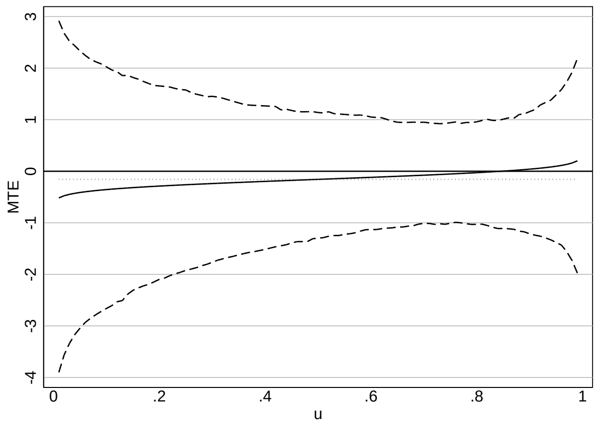
(c) *Emotional*



(d) *Hyperactivity*

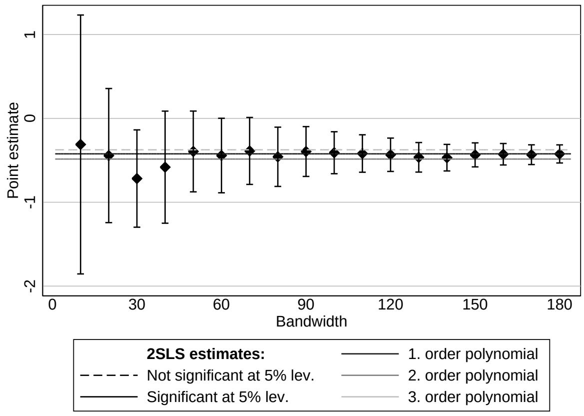


(e) *Peer problems*

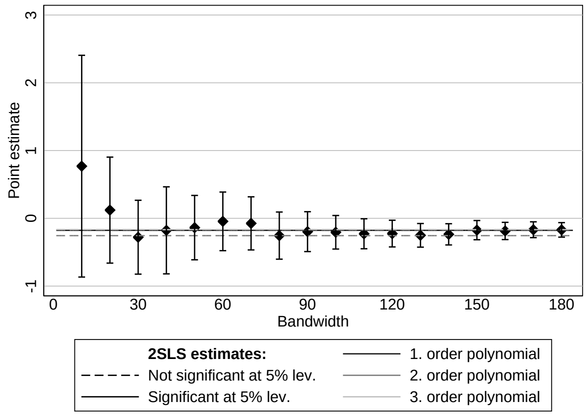


(f) *Pro-social*

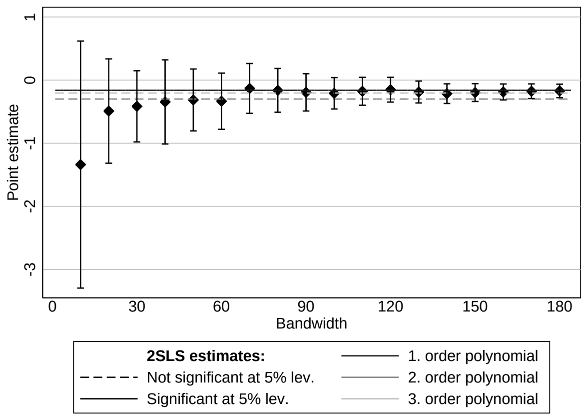
Figure 12: *Marginal treatment effects for boys, age 11. Treatment: School Starting Age > 6.5. Sample: 30 days bandwidth. We use a parametric specification in which we first estimated a probit for the propensity of being old at school enrollment, and then computed the marginal treatment effects according to Heckman et al. (2006). Bootstrapped standard errors are computed using 1,000 replications. The dotted lines indicate the average treatment effect.*



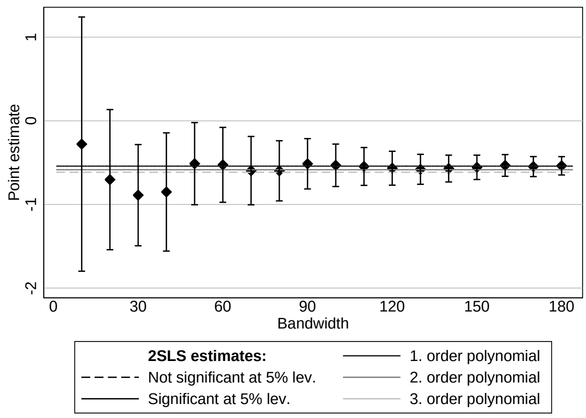
(a) *Total difficulties*



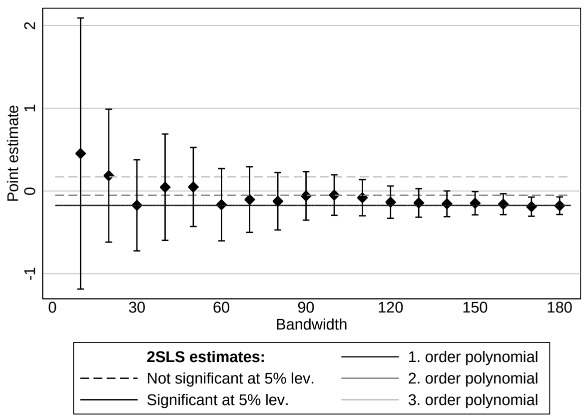
(b) *Conduct*



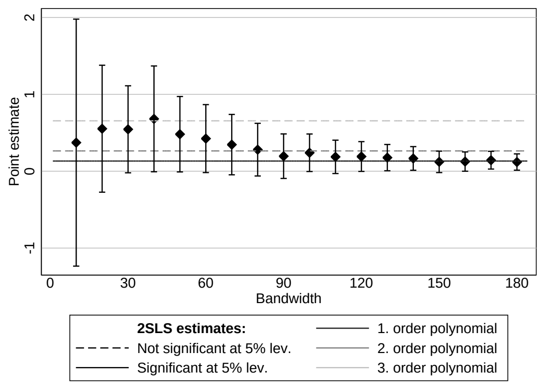
(c) *Emotional symptoms*



(d) *Hyperactivity*

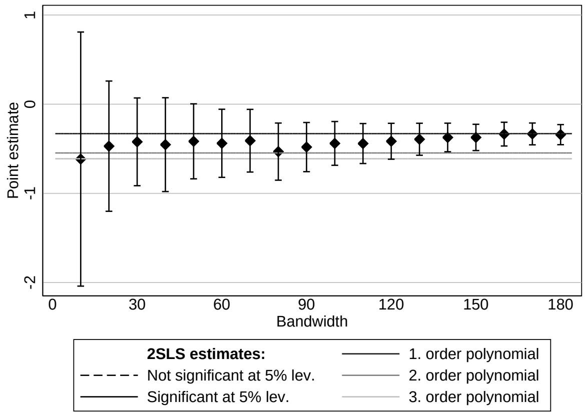


(e) *Peer problems*

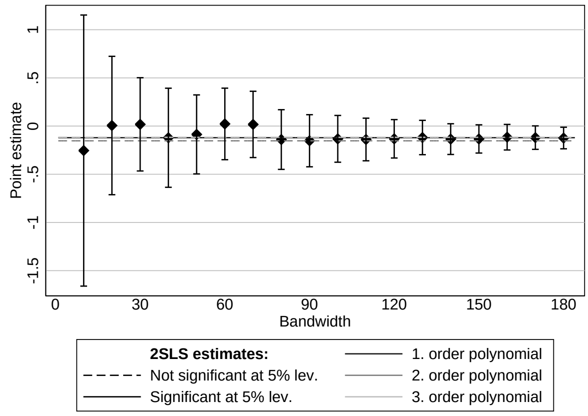


(f) *Pro-social*

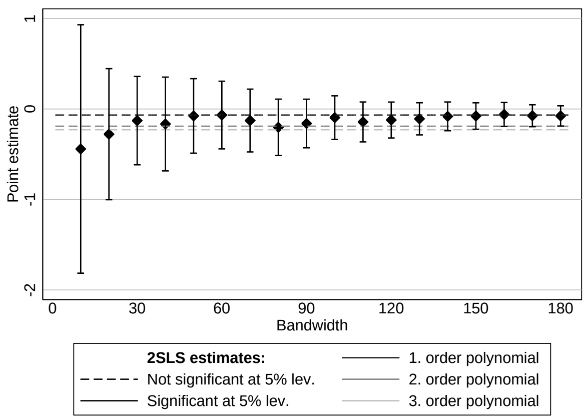
Figure 13: *Bandwidth sensitivity, age 7. Each diamond marker is the 2SLS point estimate from a local regression with the bandwidth size denoted on the x-axis. The bandwidth size increases in steps of 10 days. A bandwidth of 10 implies a sample of children born 10 days before and after January 1st. The 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.*



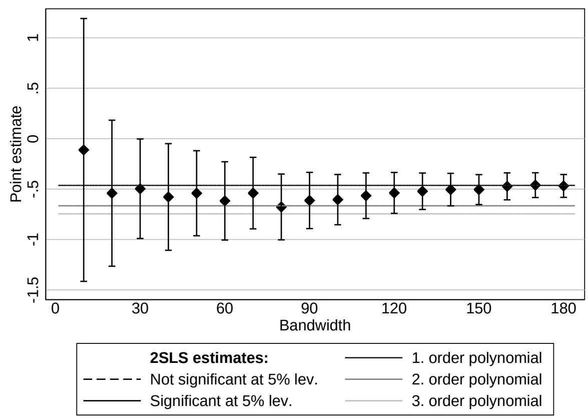
(a) *Total difficulties*



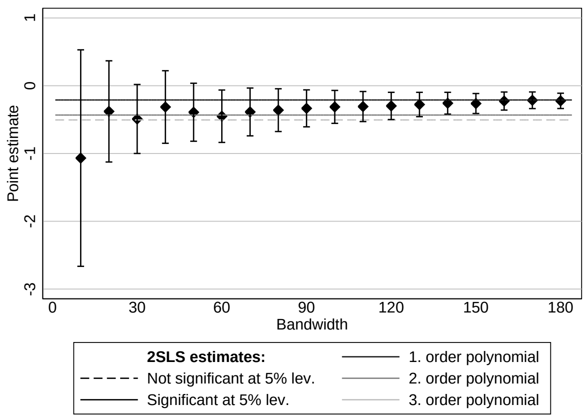
(b) *Conduct*



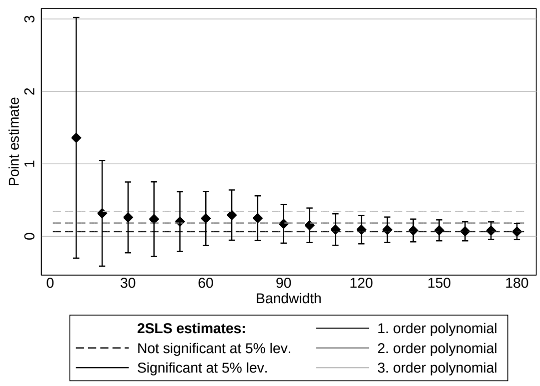
(c) *Emotional symptoms*



(d) *Hyperactivity*



(e) *Peer problems*



(f) *Pro-social*

Figure 14: *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.*

It could also be the case that other interventions are changed by the January 1st cutoff, for example changes in pre- or postnatal care programs. To rule out that such programs cause the outcomes, Figure A.10 in the Appendix shows the reduced form relationship using four outcomes measured at age 18 months. In none the cases do we find signs of jumps.

As a final robustness check we ran regressions using the survey reported school starting age, to assess the importance of a potential bias due to children retaking the kindergarten class. The results are shown in Tables A.7 and A.8 in the Appendix. The results are very similar to the main results using the imputed school starting age based on national test data.

6 Discussion and conclusions

Using data from the Danish National Birth Cohort (DNBC) linked with Danish register based data we estimate the causal effect of school starting age on non-cognitive skills in the short run. We find strong effects of school starting age on the hyperactivity scale in the Strength and Difficulties Questionnaire when the children are seven years old. Being one year older at school enrollment improves the score on the hyperactivity scale by about 0.5-0.8 of a standard deviation at age seven and 11, indicating decreased hyperactivity.

The effect is identified for compliers to the school starting rule, which states that children should enroll the calendar year they turn six. We find that compliers are considerably more likely to be girls, and that the effect is driven by girls with a low level of latent ability. For boys we only find a significant effect of school starting age on hyperactivity for the boys with the highest degree of latent ability. We find no clear evidence on the other scales of the Strength and Difficulties Questionnaire: The emotional symptoms scale, the conduct problems scale, the peer problems scale, and the pro-social behavior scale.

All children in our sample are seven years old when we measure the non-cognitive skills. Holding age constant implies that those who were older at enrollment are at the end of kindergarten class/grade zero, while those who were young at enrollment are at the end of grade one. The identified effects could therefore be driven by the fact that children are in different grades for two reasons: (1) The grades affect the non-cognitive skills differently. (2) The reference group is different. Regarding the first threat, especially for non-cognitive skills, holding age constant is likely to be more important than holding grades constant. While kindergarten class includes more element of "play and learn", both grades have class-room teaching, a centrally specified curriculum, and the same amount of minimum teaching hours. Regarding the second threat, results could be driven by the fact that those in grade zero are old compared to their cohort, and those in grade one are young compared to their cohort. However, we have shown results are robust to evaluating the subsample of children with older siblings, and to including a measure of the cohorts average age, both these robustness checks indicate that the findings are not driven by a pure reference group effect. Also, it seems unlikely that a measurement effect only affects one dimension of the SDQ.

The analysis is a based on a non-random survey of children, in which socio-economic advantaged parents are overrepresented. As compliers mainly consists of low-educated parents and effects are driven by girls with a low level of latent ability, it is likely that effects would be even stronger on a random sample of the population.

Compared to existing evidence this study is the first to use standardized measures of non-cognitive skills holding age constant. While existing evidence finds that effects of school starting age on outcomes of non-cognitive skills is driven by incapacitation, rather than human capital effects (i.e. higher human capital), our finding suggests that non-cognitive human capital indeed is affected in the short run. The fact that children who start school later have better hyperactivity outcomes may explain effects of school starting age on cognitive skills.

References

Angrist, J. D. and J.-S. Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

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.

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.

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.

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.

Datta Gupta, N. and M. Simonsen (2012). The effects of type of non-parental child care on pre-teen skills and risky behavior. *Economics Letters 116*(3), 622–625.

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., S. Urzua, and E. Vytlacil (2006). Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics 88*(3), 389–432.

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, February). 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.

Meschi, E., A. Vignoles, and A. De Coulon (2008). Parents' basic skills and childrens' cognitive outcomes. *Working Paper - London School of Economics and Political Science*.

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.

Mühlenweg, A. M. (2010). Young and innocent: international evidence on age effects within grades on victimization in elementary school. *Economics Letters 109*(3), 157–160.

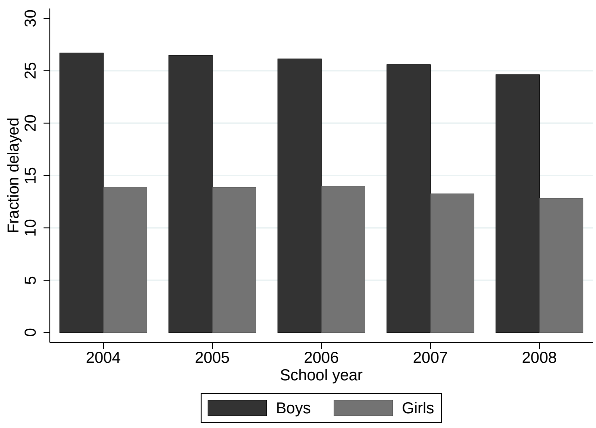
Olsen, J., M. Melbye, S. F. Olsen, T. I. Sørensen, P. Aaby, A.-M. N. Andersen, D. Taxbøl, K. D. Hansen, M. Juhl, T. B. Schow, et al. (2001). The Danish National Birth Cohort-its background, structure and aim. *Scandinavian journal of public health 29*(4), 300–307.

Rammstedt, B. and O. P. John (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality 41*(1), 203–212.

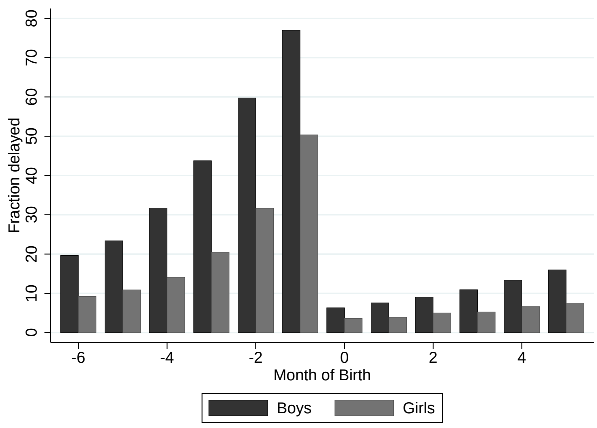
The Danish Ministry of Education (2009). Fælles mål 2009 børnehaveklassen. Technical Report 25, Undervisningsministeriets håndbogsserie.

The New Yorker (2013, September - 19). Youngest kid, smartest kid? *The New Yorker, Maria Konnikova*.

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: *Test scores in Danish and mathematics and the five dimensions of the SDQ at age 7. Girls only*

Subject	- - - Danish - - -			- - - Math - - -	
Grade	2	4	6	3	6
Emotional Symptoms 7	0.02**	0.04***	0.03***	0.01	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Conduct Problems 7	-0.06***	-0.06***	-0.06***	-0.06***	-0.07***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Hyperactivity 7	-0.18***	-0.16***	-0.15***	-0.16***	-0.13***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Peer Problems 7	-0.03***	-0.02	-0.01	-0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Pro-social Behavior 7	-0.06***	-0.05***	-0.04***	-0.02**	-0.04***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
N	9635.00	18299.00	12019.00	14908.00	11987.00

Standard errors clustered on the school level in parenthesis. $^{***}p < 0.01$ $^{**}p < 0.05$, $^{*}p < 0.1$. Each column shows results from one regression with test scores as the dependent variable, the five SDQ dimensions as independent variables and a set of covariates. Covariates included are birth weight, 5 minute APGAR score, parental education, parents' age, parental income, parental employment, mother's civil status, age at test monthly indicators (both for SDQ and the mathematics/Danish tests), school and birth year fixed effects. Both the five SDQ scores and the test scores are standardized.

Table A.2: *Test scores in Danish and mathematics and the five dimensions of the SDQ at age 7. Boys only*

Subject	- - - Danish - - -			- - - Math - - -	
Grade	2	4	6	3	6
Emotional Symptoms 7	0.03***	0.04***	0.04***	0.00	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Conduct Problems 7	-0.04***	-0.05***	-0.06***	-0.04***	-0.08***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Hyperactivity 7	-0.16***	-0.17***	-0.16***	-0.16***	-0.16***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Peer Problems 7	0.00	-0.00	0.01	0.02	0.03**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Pro-social Behavior 7	-0.04***	-0.05***	-0.03***	-0.05***	-0.04***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
N	10524.00	18763.00	11830.00	16146.00	11837.00

Standard errors clustered on the school level in parenthesis. $^{***}p < 0.01$ $^{**}p < 0.05$, $^{*}p < 0.1$. Each column shows results from one regression with test scores as the dependent variable, the five SDQ dimensions as independent variables and a set of covariates. Covariates included are birth weight, 5 minute APGAR score, parental education, parents' age, parental income, parental employment, mother's civil status, age at test monthly indicators (both for SDQ and the mathematics/Danish tests), school and birth year fixed effects. Both the five SDQ scores and the test scores are standardized.

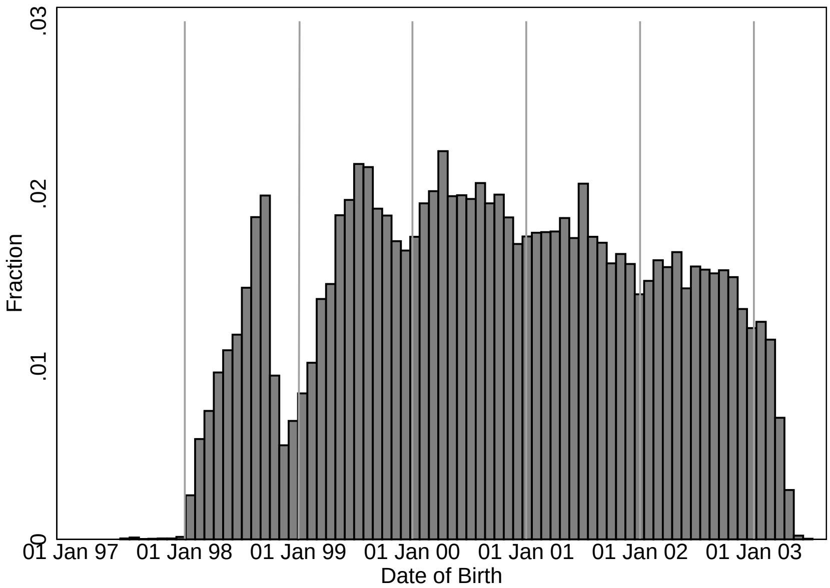


Figure A.2: *Distribution of day of birth for the 53,856 in the final analysis sample for age seven outcomes. 30day bins.*

Table A.3: *SDQ, Big Five, and Raven Score*

	- - - Big Five - - -					
	Extra-version	Agree-ableness	Conscient-iousness	Neuro-ticism	Openness	Raven
Conduct	-0.03*	0.05**	0.04**	0.01	0.01	-0.10**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)
Emotional	0.13***	-0.01	-0.05***	-0.31***	-0.04**	-0.14***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)
Hyperactivity	-0.07***	0.06***	0.18***	0.07***	0.02	-0.24***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)
Peer problems	0.15***	0.07***	-0.00	-0.01	-0.04**	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)
Pro-social	-0.09***	-0.07***	-0.08***	-0.03**	-0.07***	-0.10***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)
Mean of dep. variable	-0.00	0.00	-0.00	0.00	-0.00	8.25

Robust standard errors in parenthesis. $^{***}p < 0.01$ $^{**}p < 0.05$, $^{*}p < 0.1$ Each column shows results from one regression with test scores as the dependent variable, the five SDQ dimensions as independent variables and a set of covariates. Covariates included are birth weight (indicators for each 20th percentile) father gross income (indicators for each tenth percentile), mother gross income (indicators for each tenth percentile), an indicator for whether the father has completed higher education, an indicator for whether the mother has completed higher education. The Big Five is from the 10-item short version of the Big Five Inventory (Rammstedt and John, 2007). The Raven score consists of 12 questions. The data source is the Danish Longitudinal Survey of Children. All children are 15 years old when measured.

Table A.4: *Variable descriptives, Survey sample compared to population data, 30 days before and after the cutoff date.*

	Population data			Survey			P-value
	Mean	SD	N	Mean	SD	N	
Non-western origin	0.13	0.34	72,051	0.02	0.12	5,014	0.00
Years of schooling, highest among parents	14.30	2.59	72,051	15.51	1.93	5,014	0.00
Parents gross income	609.43	334.59	72,051	691.51	288.36	5,014	0.00
Mother's age when child was born	30.03	4.83	72,051	30.95	4.25	5,014	0.00
Father's age when child was born	32.79	5.85	72,051	33.17	5.20	5,014	0.00
Birthweight (gr.)	3477.97	618.98	73,547	3544.36	575.05	5,076	0.00
Female	0.48	0.50	73,547	0.50	0.50	5,076	0.03
5min APGAR below 7	0.01	0.08	73,547	0.01	0.09	5,076	0.19

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.

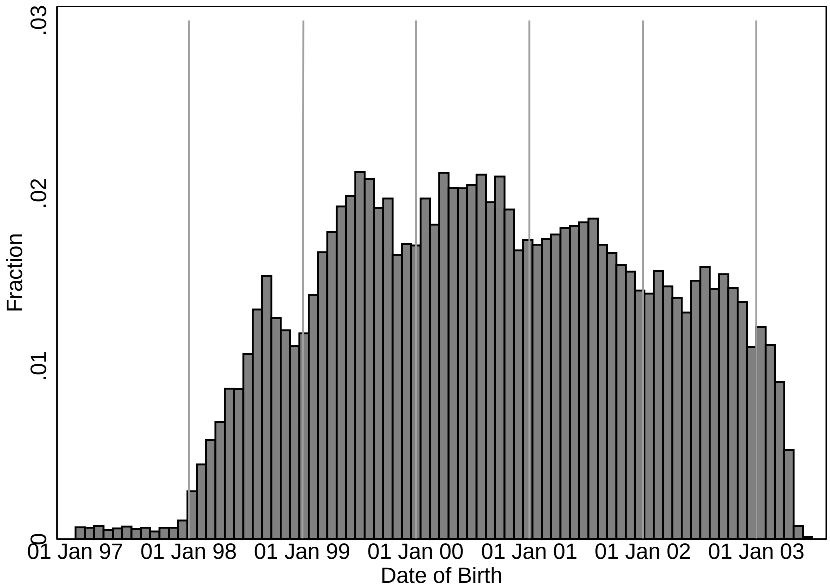


Figure A.3: *Distribution of day of birth for the 44,513 in the final analysis sample for age 11 outcomes. 30day bins.*

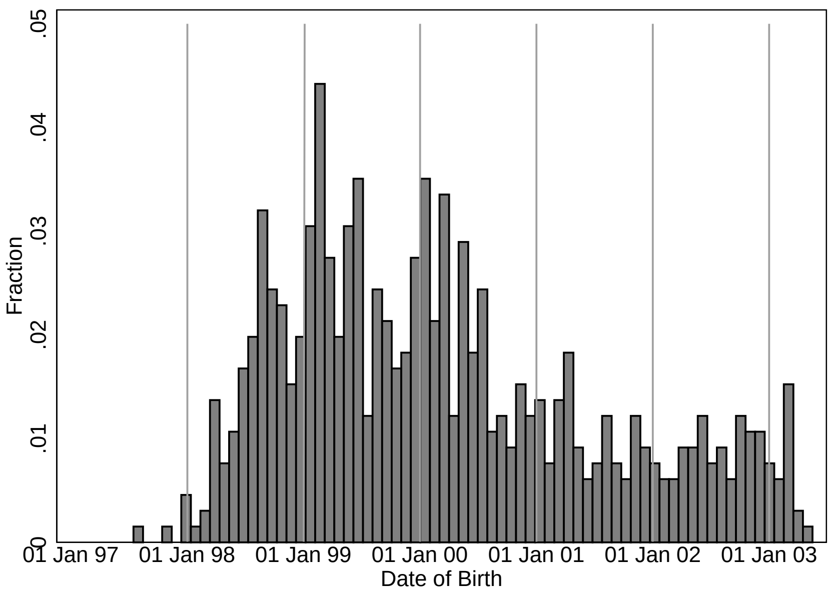


Figure A.4: *Distribution of day of birth for observations deleted due to missing administrative data. 30day bins.*

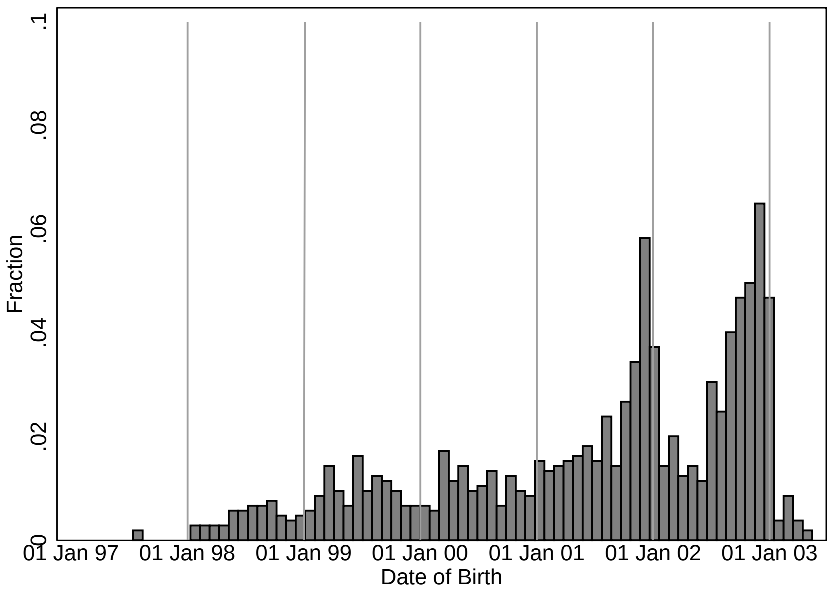
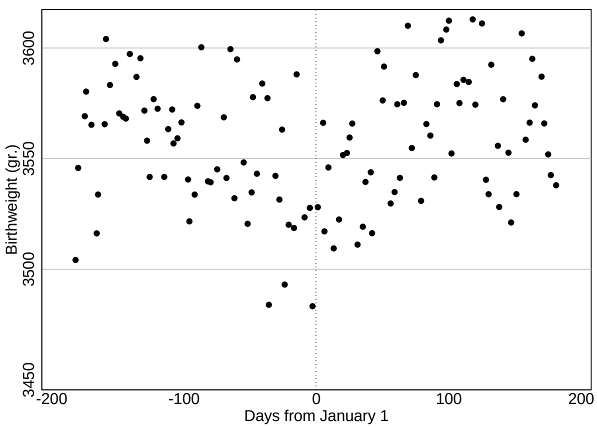
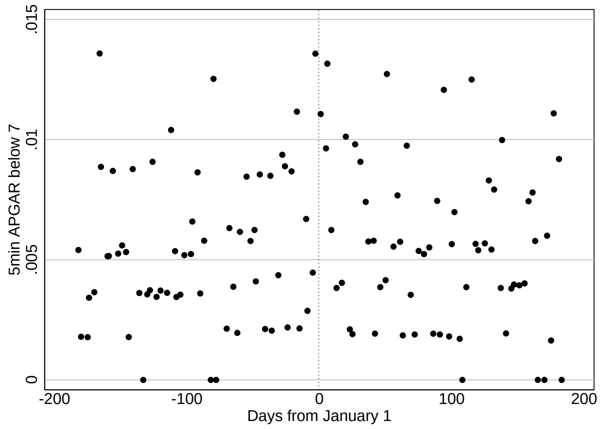


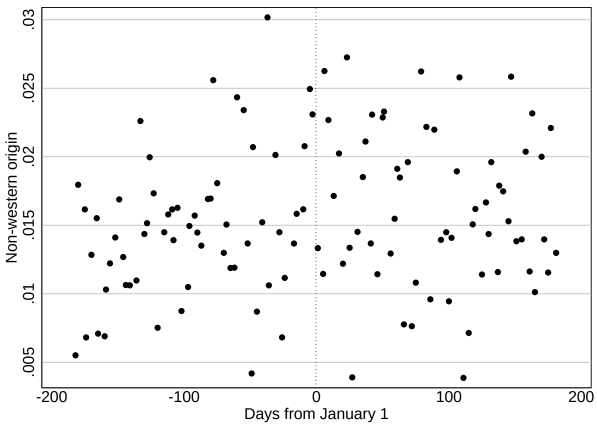
Figure A.5: *Distribution of day of birth for observations deleted due to missing parent characteristics. 30day bins.*



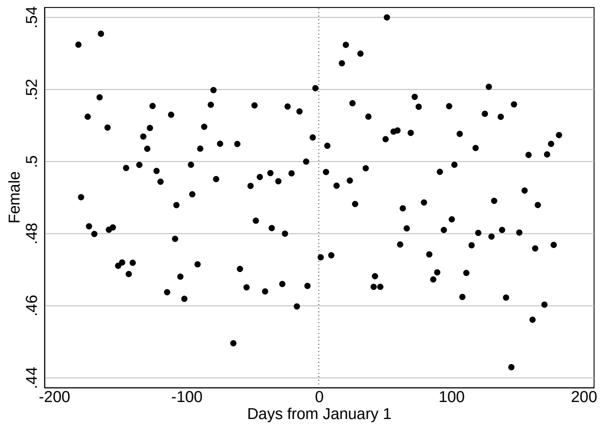
(a) *Birth weight*



(b) *5min APGAR below 7*

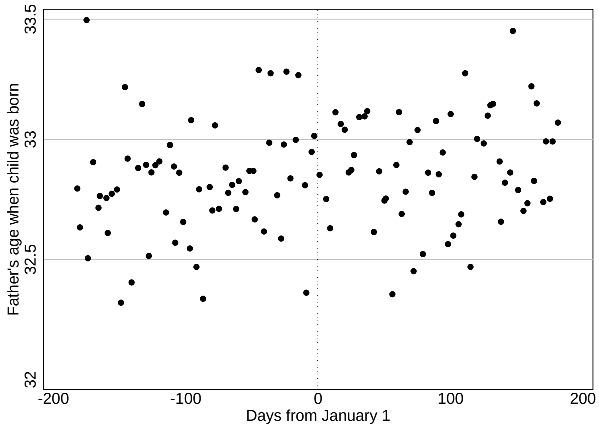


(c) *Non-western origin*

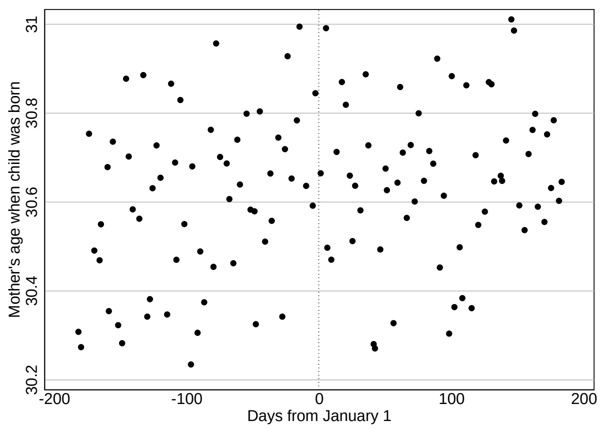


(d) *Female*

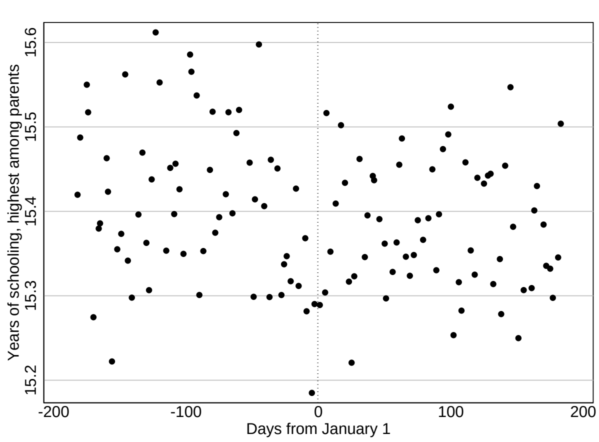
Figure A.6: *Birthday and child characteristics. Full year bandwidth & 3 day bins.*



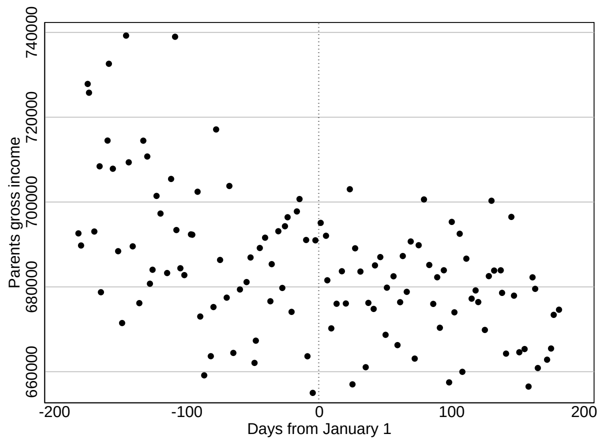
(a) *Father's age at birth*



(b) *Mother's age at birth*

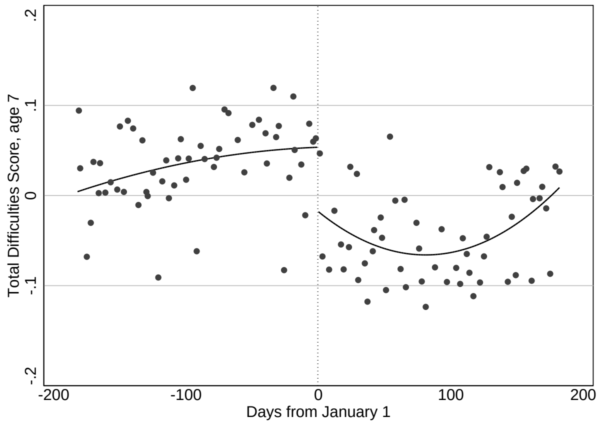


(c) *Parents' years of schooling (max)*

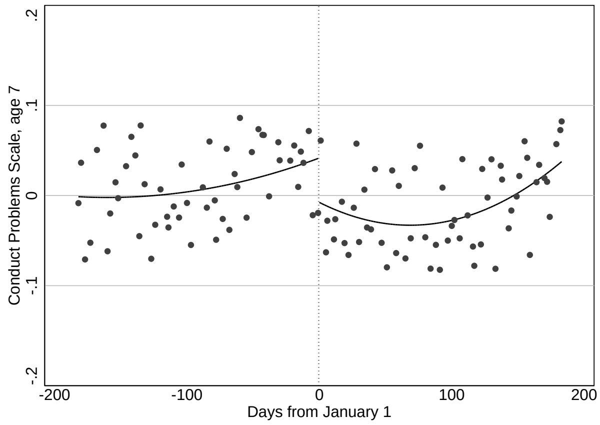


(d) *Parents' gross income (sum)*

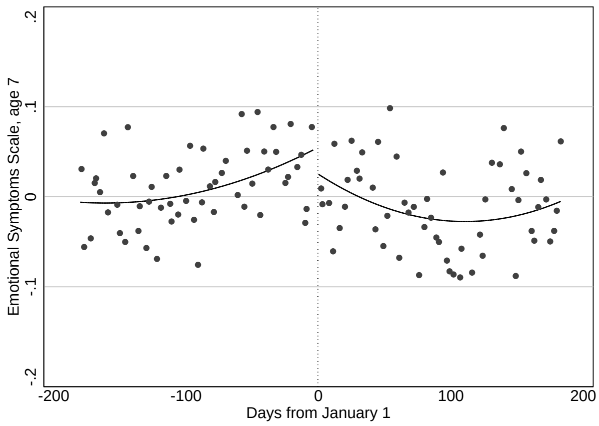
Figure A.7: *Birthday and parent characteristics. Full year bandwidth & 3 day bins.*



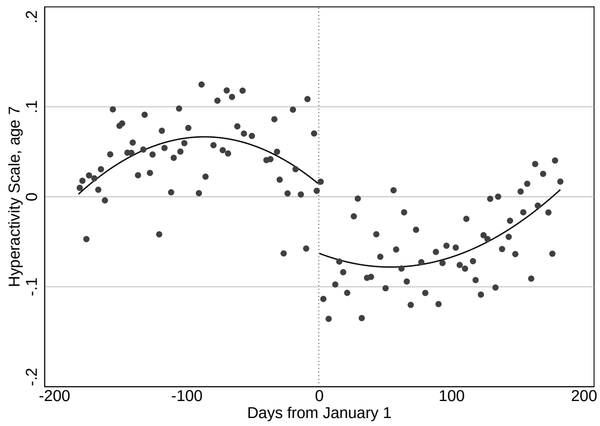
(a) *Total difficulties score*



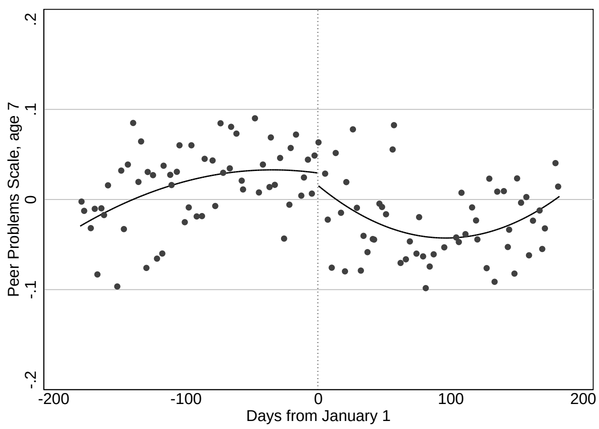
(b) *Conduct problems*



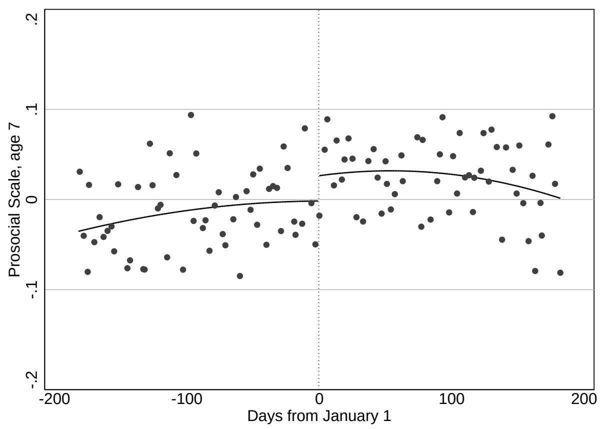
(c) *Emotional symptoms*



(d) *Hyperactivity*

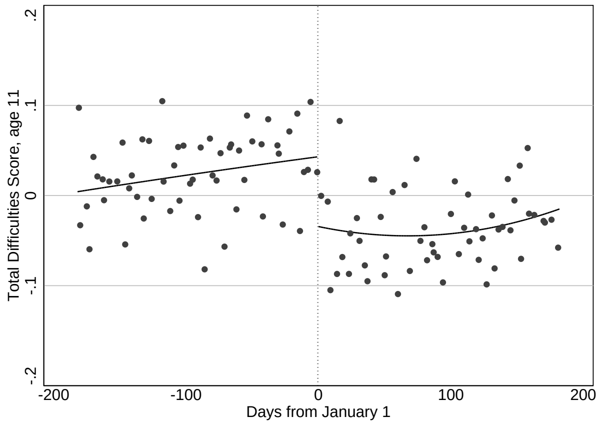


(e) *Peer problems*

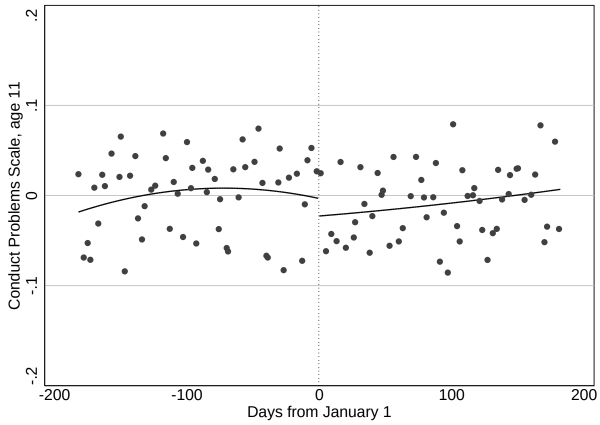


(f) *Pro-social*

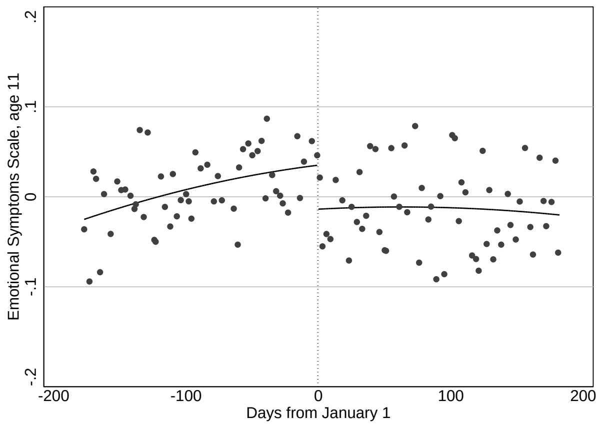
Figure A.8: *SDQ outcomes at age 7 and date of birth. All outcome variables are standardized.Full year bandwidth & 3 day bins.*



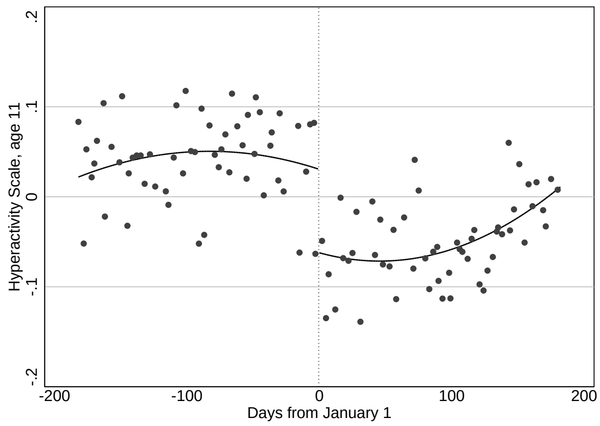
(a) *Total difficulties score*



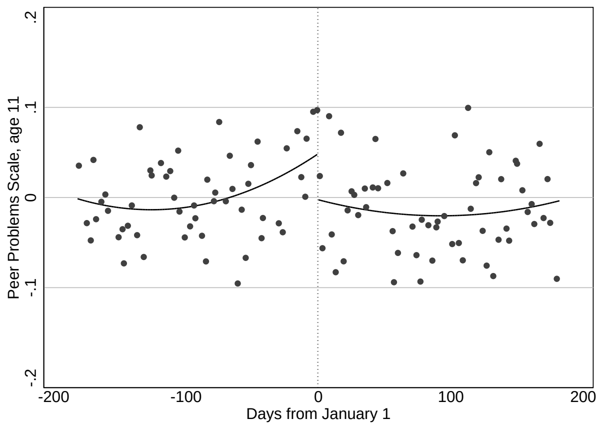
(b) *Conduct problems*



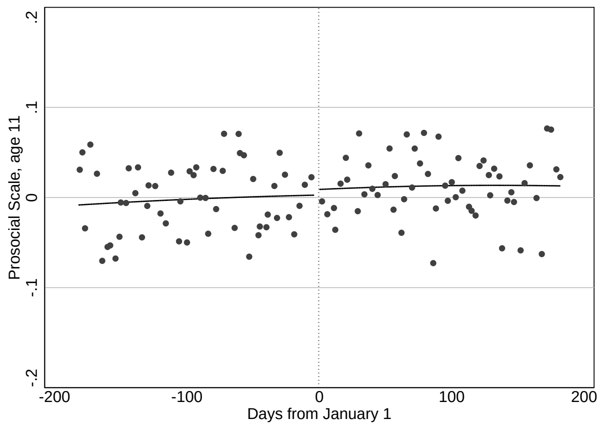
(c) *Emotional symptoms*



(d) *Hyperactivity*



(e) *Peer problems*



(f) *Pro-social*

Figure A.9: *SDQ outcomes at age 11 and date of birth. All outcome variables are standardized. Full year bandwidth & 3 day bins.*

Table A.5: *The first stage, reduced form, and 2SLS results including cohort average age, age 7*

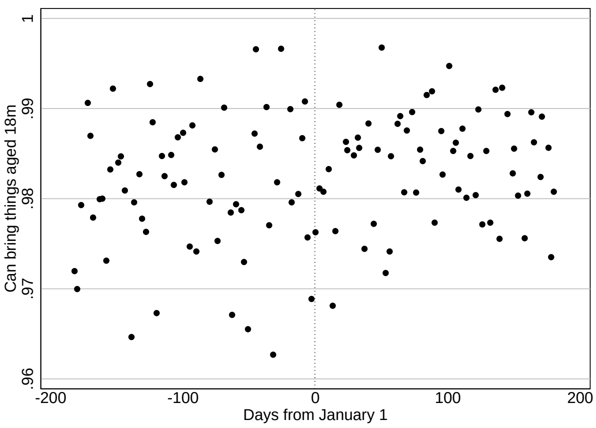
	First stage		Reduced form		OLS	2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Local specification: 30 day bandwidth & linear trends interacted with January 1. cutoff							
Birthday ≥ Jan. 1	0.17***	0.20***					
	(0.03)	(0.02)					
Total Difficulties 7			-0.12***	-0.12**	0.12***	-0.71**	-0.58**
			(0.05)	(0.05)	(0.03)	(0.30)	(0.24)
Emotional Symptoms 7			-0.07	-0.07	0.14***	-0.41	-0.33
			(0.05)	(0.05)	(0.03)	(0.29)	(0.25)
Conduct Problems 7			-0.05	-0.07	0.08***	-0.28	-0.33
			(0.05)	(0.05)	(0.03)	(0.28)	(0.25)
Hyperactivity 7			-0.14***	-0.15***	0.05	-0.87***	-0.72***
			(0.05)	(0.05)	(0.03)	(0.31)	(0.25)
Peer Problems 7			-0.03	-0.01	0.07**	-0.20	-0.04
			(0.05)	(0.05)	(0.03)	(0.29)	(0.24)
Pro-social Behavior 7			0.09**	0.07	-0.08***	0.54	0.36
			(0.05)	(0.05)	(0.03)	(0.29)	(0.24)
Observations	7,717	6,648	7,717	6,648	6,648	7,717	6,648
B. Parametric specification: Full sample & quadratic trends interacted with January 1. cutoff							
Birthday ≥ Jan. 1	0.14***	0.18***					
	(0.01)	(0.01)					
Total Difficulties 7			-0.07***	-0.07**	0.33***	-0.49**	-0.37**
			(0.03)	(0.03)	(0.02)	(0.19)	(0.16)
Emotional Symptoms 7			-0.04	-0.03	0.23***	-0.29	-0.18
			(0.03)	(0.03)	(0.02)	(0.19)	(0.17)
Conduct Problems 7			-0.04	-0.04	0.19***	-0.27	-0.26
			(0.03)	(0.03)	(0.02)	(0.19)	(0.16)
Hyperactivity 7			-0.08***	-0.08***	0.25***	-0.59***	-0.47***
			(0.03)	(0.03)	(0.02)	(0.20)	(0.16)
Peer Problems 7			-0.01	-0.00	0.25***	-0.08	-0.02
			(0.03)	(0.03)	(0.02)	(0.19)	(0.16)
Pro-social Behavior 7			0.04	0.04	-0.13***	0.26	0.23
			(0.03)	(0.03)	(0.01)	(0.19)	(0.16)
Observations	55,026	47,335	55,026	47,335	47,335	55,026	47,335
Average age of cohort		✓		✓			✓
Covariates	✓	✓	✓	✓	✓	✓	✓

Robust standard errors in parenthesis. $^{***}p < 0.01$ $^{**}p < 0.05$, $^{*}p < 0.1$. Each cell shows the estimate from a single regression. In columns (1) and (2) the dependent variable is school starting age (in years) which is regressed on an indicator for being born after January 1st, trends, and trends interacted with the January 1st cutoff. In columns (3) and (4) the SDQ measure is regressed on the same specification as in (1) and (2). Column (5) shows the results from a simple OLS regression of the dependent variable on school starting age and the time trends. Columns (6) and (7) show the 2SLS results from estimating the SDQ measure on the predicted school starting age, the time trends, and the time trends interacted with the cutoff. 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.

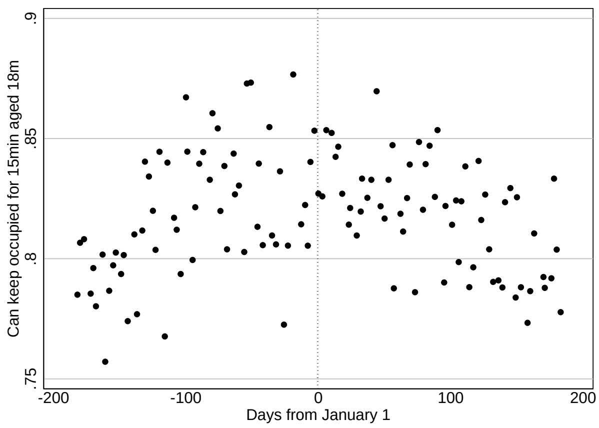
Table A.6: *The first stage, reduced form, and 2SLS results including cohort average age, age 11*

	First stage		Reduced form		OLS	2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Local specification: 30 day bandwidth & linear trends interacted with January 1. cutoff							
Birthday ≥ Jan. 1	0.20***	0.22***					
	(0.03)	(0.02)					
Total Difficulties 11			-0.08	-0.06	0.14***	-0.42	-0.26
			(0.05)	(0.05)	(0.03)	(0.25)	(0.23)
Emotional Symptoms 11			-0.03	-0.02	0.15***	-0.13	-0.10
			(0.05)	(0.05)	(0.03)	(0.25)	(0.23)
Conduct Problems 11			0.00	0.03	0.13***	0.02	0.12
			(0.05)	(0.05)	(0.03)	(0.25)	(0.23)
Hyperactivity 11			-0.10**	-0.07	0.08**	-0.50**	-0.32
			(0.05)	(0.05)	(0.03)	(0.25)	(0.23)
Peer Problems 11			-0.10	-0.07	0.04	-0.49	-0.33
			(0.05)	(0.05)	(0.03)	(0.26)	(0.24)
Pro-social Behavior 11			0.05	0.04	-0.09***	0.26	0.20
			(0.05)	(0.05)	(0.03)	(0.25)	(0.23)
Observations	6,812	6,081	6,812	6,081	6,081	6,812	6,081
B. Parametric specification: Full sample & quadratic trends interacted with January 1. cutoff							
Birthday ≥ Jan. 1	0.16***	0.18***					
	(0.02)	(0.01)					
Total Difficulties 11			-0.09***	-0.07**	0.29***	-0.55***	-0.38**
			(0.03)	(0.03)	(0.02)	(0.18)	(0.17)
Emotional Symptoms 11			-0.03	-0.02	0.23***	-0.19	-0.11
			(0.03)	(0.03)	(0.02)	(0.18)	(0.17)
Conduct Problems 11			-0.02	0.01	0.18***	-0.15	0.06
			(0.03)	(0.03)	(0.02)	(0.18)	(0.17)
Hyperactivity 11			-0.11***	-0.10***	0.23***	-0.67***	-0.56***
			(0.03)	(0.03)	(0.02)	(0.18)	(0.17)
Peer Problems 11			-0.07**	-0.06	0.17***	-0.43**	-0.32
			(0.03)	(0.03)	(0.02)	(0.19)	(0.17)
Pro-social Behavior 11			0.03	0.03	-0.08***	0.18	0.17
			(0.03)	(0.03)	(0.02)	(0.18)	(0.17)
Observations	45,630	40,428	45,630	40,428	40,428	45,630	40,428
Average age of cohort		✓		✓			✓
Covariates	✓	✓	✓	✓	✓	✓	✓

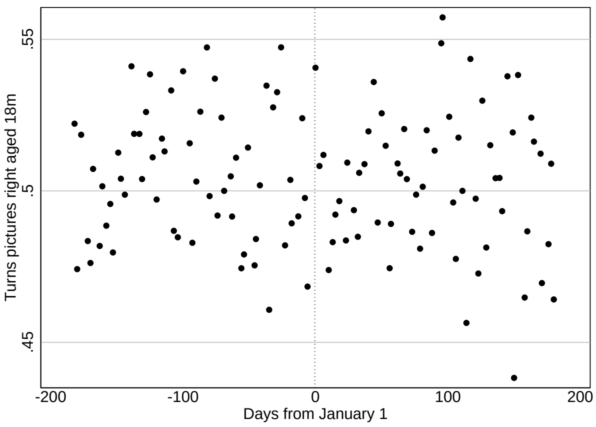
Robust standard errors in parenthesis. $^{***}p < 0.01$ $^{**}p < 0.05$, $^{*}p < 0.1$. Each cell shows the estimate from a single regression. In columns (1) and (2) the dependent variable is school starting age (in years) which is regressed on an indicator for being born after January 1st, trends, and trends interacted with the January 1st cutoff. In columns (3) and (4) the SDQ measure is regressed on the same specification as in (1) and (2). Column (5) shows the results from a simple OLS regression of the dependent variable on school starting age and the time trends. Columns (6) and (7) show the 2SLS results from estimating the SDQ measure on the predicted school starting age, the time trends, and the time trends interacted with the cutoff. 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.



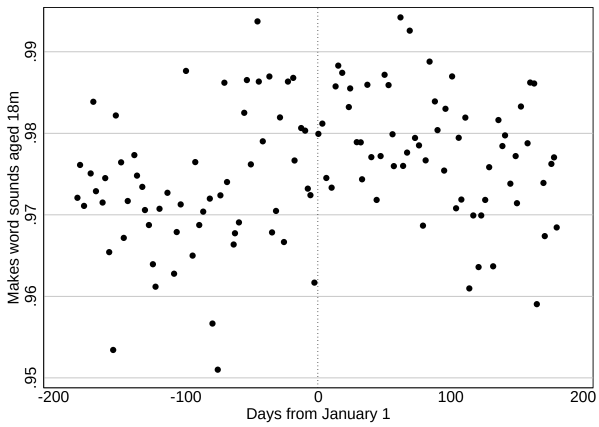
(a) *Can bring things aged*



(b) *Can keep occupied for 15min*



(c) *Turns pictures right aged*



(d) *Makes word sounds*

Figure A.10: *Child pre treatment outcomes at age 18m. Full year bandwidth & 3 day bins.*

Table A.7: *The first stage, reduced form, and 2SLS results using mother reported school starting age, age seven.*

	First stage		Reduced form		OLS	2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Local specification: 30 day bandwidth & linear trends interacted with January 1. cutoff							
Birthday ≥ Jan. 1	0.17***	0.18***					
	(0.03)	(0.03)					
Total Difficulties 7			-0.13***	-0.12***	0.10***	-0.78**	-0.67**
			(0.05)	(0.05)	(0.02)	(0.32)	(0.28)
Emotional Symptoms 7			-0.08*	-0.07	0.05***	-0.47	-0.37
			(0.05)	(0.05)	(0.02)	(0.29)	(0.27)
Conduct Problems 7			-0.05	-0.05	0.07***	-0.36	-0.29
			(0.05)	(0.05)	(0.02)	(0.28)	(0.26)
Hyperactivity 7			-0.15***	-0.14***	0.09***	-0.89***	-0.81***
			(0.05)	(0.05)	(0.02)	(0.32)	(0.28)
Peer Problems 7			-0.04	-0.03	0.06***	-0.25	-0.20
			(0.05)	(0.05)	(0.02)	(0.28)	(0.26)
Pro-social Behavior 7			0.08*	0.09**	-0.06***	0.50*	0.51*
			(0.05)	(0.05)	(0.02)	(0.29)	(0.26)
Observations	7,652	7,652	7,717	7,717	7,652	7,652	7,652
B. Parametric specification: Full sample & quadratic trends interacted with January 1. cutoff							
Birthday ≥ Jan. 1	0.14***	0.14***					
	(0.02)	(0.02)					
Total Difficulties 7			-0.08***	-0.07***	0.13***	-0.64***	-0.51***
			(0.03)	(0.03)	(0.01)	(0.22)	(0.20)
Emotional Symptoms 7			-0.05*	-0.04	0.06***	-0.38*	-0.30
			(0.03)	(0.03)	(0.01)	(0.21)	(0.19)
Conduct Problems 7			-0.04	-0.04	0.08***	-0.33	-0.27
			(0.03)	(0.03)	(0.01)	(0.21)	(0.19)
Hyperactivity 7			-0.09***	-0.08***	0.11***	-0.73***	-0.61***
			(0.03)	(0.03)	(0.01)	(0.23)	(0.20)
Peer Problems 7			-0.02	-0.01	0.11***	-0.17	-0.10
			(0.03)	(0.03)	(0.01)	(0.20)	(0.19)
Pro-social Behavior 7			0.04	0.04	-0.07***	0.28	0.26
			(0.03)	(0.03)	(0.01)	(0.21)	(0.19)
Observations	54,335	54,335	55,026	55,026	54,335	54,335	54,335
Covariates		✓		✓			✓

Robust standard errors in parenthesis. $^{***}p < 0.01$ $^{**}p < 0.05$, $^{*}p < 0.1$.. Each cell shows the estimate from a single regression. In columns (1) and (2) the dependent variable is school starting age (in years) which is regressed on an indicator for being born after January 1st, trends, and trends interacted with the January 1st cutoff. In columns (3) and (4) the SDQ measure is regressed on the same specification as in (1) and (2). Column (5) shows the results from a simple OLS regression of the dependent on variable on school starting age and the time trends. Columns (6) and (7) show the 2SLS results from estimating the SDQ measure on the predicted school starting age, the time trends, and the time trends interacted with the cutoff. 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.

Table A.8: *The first stage, reduced form, and 2SLS results using mother reported school starting age, age 11.*

	First stage		Reduced form		OLS	2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Local specification: 30 day bandwidth & linear trends interacted with January 1. cutoff							
Birthday ≥ Jan. 1	0.18***	0.19***					
	(0.04)	(0.04)					
Total Difficulties 11			-0.10**	-0.08*	0.10***	-0.61*	-0.51*
			(0.05)	(0.05)	(0.02)	(0.33)	(0.30)
Emotional Symptoms 11			-0.04	-0.03	0.05**	-0.11	-0.05
			(0.05)	(0.05)	(0.02)	(0.30)	(0.29)
Conduct Problems 11			-0.01	0.00	0.08***	-0.15	-0.06
			(0.05)	(0.05)	(0.02)	(0.31)	(0.29)
Hyperactivity 11			-0.10**	-0.10**	0.08***	-0.77**	-0.71**
			(0.05)	(0.05)	(0.02)	(0.35)	(0.32)
Peer Problems 11			-0.10**	-0.10*	0.06***	-0.57*	-0.50
			(0.05)	(0.05)	(0.02)	(0.33)	(0.31)
Pro-social Behavior 11			0.04	0.05	-0.08***	0.22	0.25
			(0.05)	(0.05)	(0.02)	(0.32)	(0.30)
Observations	5,234	5,234	6,812	6,812	5,234	5,234	5,234
B. Parametric specification: Full sample & quadratic trends interacted with January 1. cutoff							
Birthday ≥ Jan. 1	0.15***	0.16***					
	(0.02)	(0.02)					
Total Difficulties 11			-0.10***	-0.09***	0.10***	-0.61***	-0.51**
			(0.03)	(0.03)	(0.01)	(0.24)	(0.21)
Emotional Symptoms 11			-0.04	-0.03	0.05***	-0.20	-0.12
			(0.03)	(0.03)	(0.01)	(0.22)	(0.21)
Conduct Problems 11			-0.04	-0.02	0.07***	-0.34	-0.26
			(0.03)	(0.03)	(0.01)	(0.22)	(0.21)
Hyperactivity 11			-0.12***	-0.11***	0.10***	-0.76***	-0.66***
			(0.03)	(0.03)	(0.01)	(0.25)	(0.22)
Peer Problems 11			-0.08***	-0.07**	0.07***	-0.35	-0.33
			(0.03)	(0.03)	(0.01)	(0.22)	(0.21)
Pro-social Behavior 11			0.03	0.03	-0.05***	0.21	0.20
			(0.03)	(0.03)	(0.01)	(0.22)	(0.21)
Observations	37,030	37,030	45,630	45,630	37,030	37,030	37,030
Covariates		✓		✓			✓

Robust standard errors in parenthesis. $^{***}p < 0.01$ $^{**}p < 0.05$, $^{*}p < 0.1$. Each cell shows the estimate from a single regression. In columns (1) and (2) the dependent variable is school starting age (in years) which is regressed on an indicator for being born after January 1st, trends, and trends interacted with the January 1st cutoff. In columns (3) and (4) the SDQ measure is regressed on the same specification as in (1) and (2). Column (5) shows the results from a simple OLS regression of the dependent on variable on school starting age and the time trends. Columns (6) and (7) show the 2SLS results from estimating the SDQ measure on the predicted school starting age, the time trends, and the time trends interacted with the cutoff. 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.

Table A.9: *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	0.11	0.05
	(0.08)	(0.05)
Parents gross income	9814.31	-3564.30
	(13165.63)	(7784.04)
Mother's age when child was born	0.04	-0.01
	(0.18)	(0.11)
Father's age when child was born	0.20	0.03
	(0.23)	(0.13)
Birthweight (gr.)	22.25	13.45
	(25.93)	(15.09)
Female	-0.02	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. $^{*}p < 0.1$ $^{**}p < 0.05$, $^{***}p < 0.01$. 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.