CCHE CCES
Canadian Centre for Health Economics
Centre canadien en économie de la santé

Working Paper Series
Document de travail de la série

ADHD MISIDENTIFICATION IN SCHOOL: CAUSES AND MITIGATORS

Jill Furzer, Elizabeth Dhuey, Audrey Laporte

Working Paper No: 200001

www.canadiancentreforhealtheconomics.ca

January, 2020

Canadian Centre for Health Economics
Centre canadien en économie de la santé
155 College Street
Toronto, Ontario

CCHE/CCES Working Paper No. 200001
January, 2020

ADHD Misidentification in School: Causes and Mitigators

Jill Furzer*[1], Elizabeth Dhuey[1,2], Audrey Laporte[1]

[1] *Institute of Health Policy, Management and Evaluation, University of Toronto*
[2] *Department of Management, University of Toronto Scarborough*

Abstract

We estimate over- or under-identification of ADHD occurring in school-based behavioural assessments. To isolate teacher ADHD assessment error direction, we use primary school starting age and teacher-parent assessment residuals. Being young-for-grade or male generates some over-assessment. However, the under-assessment of the oldest students in a grade, especially the oldest females, drives the school starting age gap in ADHD identification. We link this gender breakdown to the growing gender gap in educational attainment. Importantly, teacher special education training mitigates these relative age-based assessment errors.

JEL Classification: I10; I12; I21; J24

Keywords: ADHD; misdiagnosis; gender gaps; human capital; teacher training

*Furzer (jill.furzer@mail.utoronto.ca), Dhuey (elizabeth.dhuey@utoronto.ca), Laporte (audrey.laporte@utoronto.ca). We sincerely thank Claire de Oliviera, Romina Tome, Borianna Milocheva, Justin Smith, Mark Stabile, Kelly Bedard, Nancy Reichman, participants from seminars at the Canadian Centre for Health Economics, the Ontario Institute for Studies in Education, the Canadian Health Economics Study Group, and the American Society for Health Economists for helpful comments to improve this paper. Jill Furzer acknowledges funding from the Canadian Research Data Centre Network (CRDCN) Emerging Scholars Grant. The analysis for this paper was conducted at the Toronto Regional Statistics Canada Research Data Centre, which is part of the CRDCN. The services and activities provided by the CRDCN are made possible by the financial or in-kind support of the SSHRC, the CIHR, the CFI, Statistics Canada, and participating universities whose support is gratefully acknowledged. The views expressed in this paper do not necessarily represent the CRDCNs or that of its partners.

1 Introduction

ADHD is the most prevalent childhood mental health disorder, in North America and internationally (Daley & Birchwood, 2010; Hinshaw et al., 2011). For students receiving special education services, almost one in four have an ADHD diagnosis (Schnoes, Reid, Wagner, & Marder, 2006; NCES, 2018). Over-identification has long been a concern due to this ubiquity and the significant growth in diagnosis over the past twenty years (Hinshaw & Scheffler, 2014). This is especially so for boys. More recently, the medical community has brought attention to the idea girls may be at risk of under-identification (Visser et al., 2014). This growth in ADHD has occurred precipitously with a rising gender gap in educational attainment, with females surpassing males in both graduation rates and higher education enrollment (Murnane, 2013). As gender differences in educational attainment have links to non-cognitive skills development (Becker, Hubbard, & Murphy, 2010), understanding gendered ADHD misidentification may help us to further understand this gap.

Ascertaining under- or over-assessment is nevertheless made difficult by limited knowledge on the true prevalence rate of ADHD (Polanczyk et al., 2014). A child's school starting age however offers an identification mechanism to address this issue, as school starting age plays a notable role in ADHD diagnosis (e.g. Elder & Lubotsky, 2009; Elder, 2010; Evans, Morrill, & Parente, 2010; Morrow et al., 2012; Schwandt & Wuppermann, 2016). A recent US study found that the youngest students in kindergarten were 1.3 times more likely to be diagnosed than their oldest peers (Layton et al., 2018). This effect is thought to be driven by teachers making relative comparisons of student behaviour within a grade (Elder, 2010). For the youngest students, teachers misattribute their relative immaturity as ADHD, while overlooking the oldest students if their behaviour is comparatively controlled (Evans et al., 2010).[1] The effect of school starting age applies not only to ADHD diagnosis, but to special education placements and disability diagnosis more generally (Dhuey & Lipscomb, 2010; Dhuey et al., in press).[2]

We use the known school starting age effect to break down over or under-identification of ADHD occurring in school and differences by gender. Our study relies on a unique data set, the National Longitudinal Survey of Child and Youth (NLSCY), that includes the Children's Behavioural Scale (CBS), an assessment tool validated to the diagnostic and statistical manual of mental health disorders (DSM). In the NLSCY, both parents and teachers were given the CBS to evaluate each individual child (Charach & Fernandez, 2013). We use the CBS within a two-stage regression discontinuity, first predicting teacher assessment errors based on parent assessments. We then isolate the direction of these errors based on school starting age. Understanding if the school starting age effect represents over or under diagnosis in ADHD addresses a long-standing unknown in the extant literature. The NLSCY also allows for evaluation of the assessment process before diagnosis to pinpoint inherent selection effects that have been left unaddressed in previous studies.

We next investigate what school and teacher characteristics drive over or under-identification of ADHD in school,

[1] The effect of a child's pre-kindergarten skills accumulation (Elder & Lubotsky, 2009) and their absolute age at the time of behavioural assessment (Black, Devereux, & Salvanes, 2011; Cascio & Schanzenbach, 2016) on actual witnessed behaviour do not go unacknowledged in this paper. We, however, focus on how a child's placement within the age distribution may lead to a faulty referral by not conditioning a behavioural assessment on a child's absolute age.

[2] There is an extensive literature on school starting age effects and their effect on a myriad of outcomes. See Dhuey et al. (in press) for a review of the literature.

knowing the impact of school quality on the gender gap in educational attainment (Autor et al., 2016). School factors include peer behaviour, and school-wide socioeconomic levels and disability prevalence. We also focus on teacher experience, age and their special education training, given the importance of teacher quality to overall student success (Hanushek, 2011). To the best of our knowledge, we are the first to investigate these plausible mechanisms of both the school starting age effect, but also within the context of gendered differences in education settings.

Identifying the type of ADHD misidentification occurring is crucial for understanding how to remediate its effects on educational outcomes. Given the overlap of ADHD and non-cognitive skills, missed or late diagnosis constitutes missed human capital investment if a child does not receive necessary early educational or medical interventions (Doyle et al., 2009). Timing interventions to sensitive developmental periods is key to ensuring returns to education investments (Heckman, 2007; Lubotsky & Kaestner, 2016). On the other hand, over-diagnosis amounts to misallocation of school and medical resources and presents individual student costs, including adverse impacts on schooling performance from treatment via stimulant medication (Currie, Stabile, & Jones, 2014; Chorniy & Kitashima, 2016). Stigma from peers, family and school staff (Moses, 2010; Bharadwaj, Pai, & Suziedelyte, 2017) and negative expectations from teachers (Ohan et al., 2011) are also of concern. If these adverse outcomes fall disproportionately to one gender or the other, there is a definite cause for assessment errors contributing to the overall gender-education gap. The effects of appropriate identification may also extend beyond the individual child, by mitigating negative externalities in the classroom (Brown et al., 2001; Pelham et al., 2000; Aizer, 2008). Decomposing the direction of ADHD assessment error leading to misdiagnosis is thus vital for targeting interventions to alleviate this potential misallocation.

Our results first show that ADHD assessment errors in early grades based on school starting age originate with teacher assessments. We find little evidence of school starting effects in parent assessments. Using these results, we estimate residual teacher assessments, conditioning on parent assessments to remove underlying correlation. We then regress this residual teacher assessment on school starting age to decompose under- or over-assessment of ADHD severity contained within the school starting age effect. This method gives a similar result to a simple regression discontinuity on teacher assessments, controlling for parent assessments while also centering the effect around a mean of zero. The assessment error direction is then evaluated based on deviations above or below zero.

We find a robust negative deviation in assessments for the oldest students within a grade, which we interpret as an under-assessment. We also see positive deviations in assessment error for the youngest in a grade, though we calculate that under-assessment for the oldest students constitutes 60% of the total school starting age effect. Regardless, when broken down by gender, we find that under-reporting of ADHD symptoms for the oldest females drives the school starting age effect in full. Teacher assessments of ADHD in females also under-report symptoms across the relative age distribution. At the same time, teachers over-report symptoms for males, irrespective of school starting age. These results align with the growing medical consensus that ADHD is often missed in girls (Quinn & Madhoo, 2014) and runs parallel to identified gendered differences in mental health assessments found in countries with low ADHD prevalence rates (Dee & Sievertsen, 2018). While the extant literature shows minimal gender differences in ADHD diagnosis by school starting age in countries with high diagnosis rates (e.g. Evans et al., 2010), this is likely because the majority of students referred for diagnosis in these countries are boys. In identifying these results, we are the first to elucidate pre-diagnosis teacher ADHD assessment errors and to address the direction of these errors, leading to gendered selection effects for eventual diagnosis.

In investigating mechanisms of over or under assessment, we find that the proportion of low-income students in a school increases assessment errors. This result holds even when controlling for the student's own family income. General classroom behaviour also reflects highly in the teacher's assessment of individual student behaviour, especially so for boys. Together, these results offer a mechanism whereby school quality impacts educational attainment and the gendered gaps therein. Being in a school with high levels of disadvantage, with corresponding effects on the non-cognitive skills of peers (Autor et al., 2019), generates an assessment error for the individual student. Nevertheless, we find teacher education, specifically special education training or designations, plays a mitigating role against assessment errors. Beyond the study of assessment errors in ADHD, teacher special education training may act as a potential solution to the widespread issue of adverse school starting age effects. There are plausible extensions to special education tracking, learning disorder diagnosis and general educational attainment, especially when these occur more so for males.

As a final contribution, we find that the internal or external nature of the symptoms for ADHD inattentive and hyperactive sub-types serves to impact the direction and magnitude of assessment errors. This effect runs concomitantly with gender and translates to other co-morbid disorders that have both internal and external presenting symptoms such as anxiety, depression and conduct disorder (Jensen, Martin, & Cantwell, 1997). We discover stronger under-assessment of hyperactive symptoms for the oldest females in a grade versus inattentive symptoms. For ADHD co-morbidities, as conduct disorders are more prevalent in males, they are also more likely to be over-assessed, regardless of school starting age, but especially so for the youngest males. The oldest females are under-assessed for conduct disorder but display stronger school starting age effects in anxiety or depression.

In presenting these results, the following section provides an overview of the in-school referral process leading to assessment errors in ADHD. Section three reviews our data set, the Canadian NLSCY, and details our identification strategy based on school starting age. Results are presented in section four, covering both school starting age estimates, calculation of assessment error direction and the mechanisms of this effect. We conclude in section five.

2 In-school ADHD assessment errors

Identifying young children with emotional and behavioural disorders requires multiple levels of referral and agreement between physicians, families, and teachers, across often conflicting incentives (Edmunds & Martsch-Litt, 2008). Teachers and administrators significantly influence this diagnostic process, with 46% of diagnoses initiated by teachers and 6% by schools (Sax & Kautz, 2003). The DSM also includes impairment in school as a criterion for diagnosis and multiple questions specifically related to classroom behaviour (see appendix A for full DSM criteria, e.g., the student makes careless mistakes in schoolwork) (American Psychiatric Association, 1994). Moreover, over 80% of primary care physicians report using school information, like behavioural ratings made by teachers and school psychologists, and report cards, in making their ultimate diagnostic decisions (Chan et al., 2005).

Subjectivity is, however, inherent in this diagnostic process as school psychologists and physicians rely on the DSM to guide what separates ADHD from normal levels of behaviour (Langberg, Froehlich, Loren, Martin, & Epstein, 2008). The DSM suggests a threshold of six or more of the specified 18 ADHD symptoms for diagnosis, with nine symptoms each for hyperactivity and inattentiveness. These behaviours often mirror normal child misbehaviour but are required to be sustained and impacting upon daily activities. Bias in diagnosing ADHD, whereby males are more likely to be diagnosed, all else equal, has been found previously in physician decision making (Bruchmüller et al., 2012). In fact, ADHD is specifically prone to subjectivity based on its high prevalence and potential for stereotypes (Ohan & Visser, 2009). With a high male to female diagnosis ratio, boys have become the prototypical ADHD patient (Faraone et al., 2003). Though there is a biological component to ADHD rendering it more likely to manifest in males, ADHD is often viewed as being a male disorder, which may drive gender-based differences in assessment (Quinn & Madhoo, 2014). ADHD is also a spectrum disorder defined by three subtypes: predominantly inattentive, predominantly hyperactive-impulsive, and a combination of these two types (American Psychiatric Association, 1994). As students with externally presenting symptoms like hyperactivity-impulsivity negatively impact the classroom learning environment of their peers (Aizer, 2008; Figlio, 2007; Fletcher, 2010), hyperactivity-impulsivity symptoms may be over-weighted in referral decisions. Reciprocally, the limited salience of more internalized inattentiveness may produce under- or late identification.

Gender and symptom-type factors are also likely interactive. While hyperactive-ADHD is more common in males, females are more likely to have inattentive-spectrum ADHD, without visible hyperactivity (Gaub & Carlson, 1997; Gershon, 2002). Taken together, males with any hyperactive or impulsive behaviours are already at risk of being over-referred for ADHD diagnosis. Students with inattentive symptoms, female students, or those without visible school impairments may then be overlooked. These hypotheses are tantamount to representative typing in decision-making (Kahneman & Tversky, 1972). Relative age within a grade may exacerbate this further.

When teachers evaluate student behaviour, younger than average students start kindergarten with a less developed set of non-cognitive skills (Elder & Lubotsky, 2009). Males are also less likely to be kindergarten ready (Autor et al., 2016). Thus, while age-appropriate, the behaviour of younger students or males in a grade may present as disruptive in comparison to older peers. If this relative misbehaviour is viewed by the teacher as disordered, any referral initiated is based on an initial misattribution. Conversely, a child with the same level of underlying ADHD but who waited a year to enroll in primary school is less likely to be referred for assessments if their relatively higher levels of self-regulation allow them to control or mask their ADHD symptoms.

Even if a teacher considers a student's absolute age or gender when assessing behaviour, teachers might still need to contend with how a student's behaviour impacts the learning of their peers. If a student within a grade takes up significant time or is overly disruptive to the classroom environment, the teacher might seek accommodations or assistance for this child to alleviate classroom constraints. This is still a referral pathway driven by necessity, but not having to do with an underlying disorder. The known diagnostic error based on school starting age is thus driven first by a selection in referral decisions made by the teacher. In early schooling years, a child at-the-margin of confirmable

ADHD has a higher likelihood of referral than one whose birthday falls after the cut-off date. Skill differences by absolute age will also persist well past school start, leading to cumulative long-term effects (Black et al., 2011; Cascio & Schanzenbach, 2016).

Any formal assessments by a school psychologist or an outside mental health expert are then performed on a subset of students referred based on their relative maturity, gender or symptom type. At this point, school psychologists or physicians do play some role in diagnostic "gate-keeping", with evidence of higher special needs referral dismissals for the youngest students (Balestra, Eugster, & Liebert, 2017). However, previous work shows that younger students are still much more likely to be diagnosed with ADHD than their older peers, 5.4 percentage points more likely, by some estimates (Elder, 2010). Younger students also receive more intensive medical treatment for ADHD (Layton et al., 2018). A teacher's initial assessment of a child's relative behaviour thus has a significant bearing on which students at the margin ultimately receive a diagnosis. Identification of teacher-based over and under ADHD assessments is critical for understanding its role in educational attainment gaps overall. As ADHD is a deficit in non-cognitive skills (Currie & Stabile, 2006), and non-cognitive differences generate a gendered gap in education (Jacob, 2002; Becker et al., 2010), these two things are inextricably linked.

It follows from this that numerous school, class and teacher factors may serve to drive demand for diagnoses of the youngest students, and thus assessment error. Previous work has identified that supply-side physician factors attenuate diagnostic error based on school starting age effects through specialist "gate-keeping" in low-prevalence countries (Dalsgaard, Humlum, Nielsen, & Simonsen, 2012). In the school, both class size and the share of foreign-born students in a classroom may increase assessment error to some degree (Schwandt & Wuppermann, 2016). Knowing the pivotal role of teachers and school quality in a student's success, exploring other school-based demand-side factors is warranted. The extensive work on teacher-value add shows us that teacher experience plays an active role in later life outcomes (Hanushek, 2011). Mitigating a assessment errors may be a plausible channel through which this experience factor works. Furthermore, teacher special education training has been shown to affect outcomes for students with disabilities positively (Feng & Sass, 2013). The theoretical impact of school-wide disability burdens and level of disadvantage may also serve to constrain school resources, leading to higher assessment errors.

3 Data and Methods

3.1 Data

We rely on Statistics Canada's National Longitudinal Survey of Child and Youth (NLSCY) to investigate under and over assessment in ADHD identification. The NLSCY was initiated in 1994 with follow-up surveys every two years until 2008. It covers all ten Canadian provinces, excluding children living on federally designated First Nations Reservations, in remote northern locations and the Canadian Territories. A person most knowledgeable, which was the biological mother in 90% of cases, responded to questions regarding their child and household, including personal background information about health and education of both parents if applicable, and information around the child's health, education, and behaviour (Statistics Canada, 1996). Out to 2002, the NLSCY also included comprehensive surveys from the child's school. These teacher surveys asked the teacher to report on the student's school performance and behaviour, that of their peers, and to provide personal demographics, training and experience information. This survey provides us with a comprehensive set of family, teacher, and school factors to both precisely estimate assessment error and to explore individual-level mechanisms behind these errors.

The NLSCY is also distinct through its inclusion of the Children's Behavioural Scale (CBS) for inattentiveness/hyperactivity. The CBS is a behaviour and emotion assessment tool validated against the then-current diagnostic and statistical manual of mental health disorders (DSM-III) (Charach, Lin, & To, 2010). The inattentiveness/hyperactivity scale consists of eight symptom components that correspond to the DSM-III, each assessed on levels of severity from none, sometimes or always. The full list of questions is found in Appendix A. Both a parent and the child's teacher separately assessed the child's behaviour using this scale. If a parent or teacher reported that a child never experienced a given symptom, the child was given a corresponding score of 0. Sometimes experiencing a symptom resulted in a score of 1 and "always" a score of 2. The highest score a child could have in the separate parent and teacher assessments was 16. The lowest possible score was 0. The average teacher ADHD assessment score was 3.50 (3.69) and the average parent assessment was 4.49 (3.17) (Table 1).

As the NLSCY consists of both longitudinal and cross-sectional cohorts, we pool across all waves corresponding to years 1994-1995 to 2002-2003 and select all children enrolled in kindergarten at the time of the survey. This guarantees that children only appear once. Selecting only kindergarten also ensures no selection into specific teachers or educational streams based on prior knowledge of a child's in-school behaviour or ability. Kindergarten-aged children were additionally over-sampled in the NLSCY, leading to a larger set from which to work. We include only cycles one to five as the NLSCY included surveys from teachers and schools only during these years.

One caveat in using teacher and school variables is that only 48 to 61 percent of teachers responded to NLSCY surveys, depending on the cycle year (Statistics Canada, 2002). The most significant reason for failure to report was due to teacher non-response or administrative error, with no calculated response bias. Non-response due to withholding of consent did occur; this was highest in cycle one with 3.4% of parents refusing consent and 9.6% of school boards (Statistics Canada, 1997). An analysis of non-consenting individuals suggested some evidence that children from lower income families and who were reported to be doing poorly in school were more likely to be non-respondents. For the remaining sample, average teacher assessments of child misbehaviour might be lower than what would be expected for the population given an association between income and ADHD (Currie & Stabile, 2006). With the potential exclusion of children with more behavioural issues, any realized results can be interpreted as lower bound estimates.

To calculate age at school start, we use the mandated eligibility date for which a child could enroll in kindergarten in each province. In the provinces of Alberta and Saskatchewan, eligibility cut-off dates for enrollment in kindergarten are decided at the school-board level. In instances where either school-board eligibility dates could not be established or where a city could not be linked to a specific board, the observations were dropped. This exclusion resulted in the elimination of approximately 1,300 observations from the analysis. There are an additional 35 observations where there are no parent-reported responses to the ADHD behavioural index when there was a teacher assessment available. We further excluded any child who is not assigned a specific Canadian home city and Province (42 observations). The teacher-reported average ADHD behavioural index for these exclusions is larger at 5.82 vs. 3.90. This is however driven by several outliers. Following these exclusions, the total sample size is 7,510 children over five cycles. We use a secondary sample of students from kindergarten thru to grade six for supplementary analysis that contains 20,315 students.

3.2 Empirical Strategy

To measure assessment errors using age at school entry, our preferred model specification is a non-parametric regression discontinuity design based on the distance between the eligibility cut-off date for school entrance in school board c and the birth date of a child i: $days_i = birthdate_i - eligibilitydate_c$. This is the days between a child's birth date and the specific province or school board cut-off date. For instance, $days_i$ is equal to 0 for children born on a cut-off date, 1 for those born the day after and -1 for those born the day before. The cut-off date is most commonly December 31^{st} in Canada, but ranges from September 30^{th} to March 1^{st} across provinces. A full list of cut-off dates across provinces and years can be found in Appendix B. This running variable is centred around zero such that those that are eligible for enrolling in a given year will have $days_i \leq 0$ and those ineligible will have have $days_i > 0$. Estimating the effect of a being young for one's grade on teacher perception of their ADHD severity is then modeled as:

$$Y_i = \alpha_1 + \beta_1 Young_i + g(days_i) + X_i^{'}\gamma_1 + \mu_i \quad (1)$$

The outcome variable Y_i represents a teacher's ADHD assessment of student i. We quantify if a child is young for grade by calculating the child's age on the first day of school, assumed to be September 1^{st}, using the child's exact birth date. We then compare the child's age at school start to the median age for children in their grade, province, and cycle year. If their age is less than this median age, a value of one is assigned to the indicator variable for $Young_i$. The coefficient α_1 is thus our treatment effect of interest, measuring the impact of being young-for-grade at the discontinuity point on teacher assessments. This can be interpreted as the treatment effect of enrolling in school one year later, given that the regression is centered around distance between eligibility cut-off date and birth date.

We include the operator $g(.)$, which is a smoothed polynomial function of $days_i$, allowing for a non-linear construction of ADHD assessments by the running variable, on either side of the cut-point. In practice, our final results are obtained from a local-linear model estimated on either side of the cut-point for individuals falling within a local bandwidth of $days_i$ observations, knowing limitations of higher order global polynomials in regression discontinuities (Gelman & Imbens, 2019). While not required for identification, the matrix X_i contains gender, year of assessment, the log of total household income, mother's age, the number of children in household and regional fixed effects to improve precision of estimates. Regional fixed effects correspond to Canada's four eastern maritime provinces, the four western provinces and separate indicators for the two most populous provinces of Quebec and Ontario. Provincial fixed effects are not used based on sample size restrictions in smaller provinces.

It is likely that $days_{ic} \leq 0$ will not perfectly predict if a child is young for their grade. Children can be held back from entering school when first eligible and in a small number of cases, parents can petition to let their child start school earlier than the eligibility date (Bassok & Reardon, 2013). If underlying ADHD-like behaviours act as selective mechanisms causing parents to either enroll their child on time, early, or late, then our estimates of the school starting age effect will be biased. Despite having a kindergarten sample and thus knowing exact age at school enrolment, to control for selection, we use a fuzzy RD. This relies on an estimation of the same local linear function on either side of a province-specific kindergarten eligibility cut-off to first predict whether a child will be young for their grade (Imbens & Lemieux, 2008). This is then used to instrument our above equation (1). A fuzzy RD has been shown to be similar to a two-stage least squares approach (Hahn, Todd, & Klaauw, 2001). Likelihood of a child being young-for-grade is based on their age relative to the school board or provincial cut-off date:

$$Young_i = \alpha_2 + \beta_2 I(days_i > 0) + g(days_i) + X_i^{'}\gamma_2 + \upsilon_i \tag{2}$$

We then estimate the effect of being predicted young for one's grade on the likelihood of ADHD identification in a second stage. This generates a local average treatment effect (LATE). We expect the LATE to be positive: a child who is predicted to be young for their grade will have higher teacher-perceived ADHD issues. If we run equation (1) using only those who were compliant with the eligibility cut-off date by excluding any early or late enrolled students, we expect a smaller treatment effect that is more precisely estimated. We expect comparator regressions using parent-reported scores to be close to zero regardless of specification.

Using school starting age as an identification strategy has been criticized for its potential to violate the monotonicity assumption required for any instrumental variable (IV) design, or IV-equivalent like a fuzzy RD (Barua & Lang, 2016). The closer a child's birthday is to the cut-off date, the higher the likelihood of a parent holding their child back might be. This would decrease their likelihood of being young for a grade and result in non-monotonicity. This criticism is limited for data sets such as the NLSCY, which are able to precisely measure effects at the cut-point, based on individual-level birth dates over using month or quarter of birth approximations. In our kindergarten sample, we also know exact school starting age. We can still, however, employ several robustness checks to test for this possibility and whether the estimated effect remains consistent. We use extensive control variables and test heterogeneity in red-shirting behaviour of parents, as proposed by Dhuey et al. (in press). We also employ regional fixed effects. Given some province-specific differences in allowance of red-shirting or early enrollment, notably in Quebec, inclusion of these effects will account for unobserved norms in parental behaviour that may undermine the monotonicity assumption.

Our results utilizing a fuzzy RD design would also be compromised if children on either side of the cut-point were dissimilar or if there were non-random sorting into age at school entry outside of red-shirting. While there are some instances of birth timing (Buckles & Hungerman, 2013) and seasonality of some disorders based on birth timing (Currie & Schwandt, 2013), evidence of timing birth to a school cut-off date is limited (Dickert-Conlin & Elder, 2010). ADHD itself also does not appear to be linked to the season of birth (Schwandt & Wuppermann, 2016). However, given the association of birth timing with family socioeconomic status, we use a wide range of controls on family income, maternal age and family structure to remove observable heterogeneity leading to birth timing. Finally, our data sample employs a variety of eligibility cut-off dates across all provinces in Canada and thus seasonality overall should be negligible. To test for any remaining seasonality, we run placebo tests for asthma diagnosis, another common childhood disorder with known seasonality and find no effect by school starting age (Appendix Table A5).

3.2.1 Assigning Direction of Misdiagnosis

To break down the magnitude of over-identification and/or under-identification of ADHD driven by school starting age, we use both teacher and parent-reported scores. We posit that a regression of teacher-reported scores on parent-reported scores will produce residuals consistent with the measurement error in teacher assessment, conditional on parent assessments. If measurement error is assumed to be random, these are normally distributed around mean zero.[3] As we expect teacher assessments of ADHD to jump arbitrarily at the eligibility cutoff date for kindergarten, their assessment error (ξ_i^T) should have non-zero heterogeneity around these same cut-offs. Based on previous findings and as shown in our results section, school starting age effects are not present in parent assessments of their kindergarten-aged child (Elder, 2010). We thus expect parent assessment error (ξ^P) to have no relative age-related heterogeneity. To precisely measure the direction of teacher assessments based on schools tarting age, we estimate a new equation according to:

$$\eta_i = \alpha_3 + \beta_3 Y\hat{oung}_i + g(days_i) + X_i^{'}\gamma_3 + v_i \tag{3}$$

Here, η_i is the residual teacher assessment error. All other parameters are analogous to equation (1). We then assess direction knowing $\beta_3 = \tau^Y - \tau^O$. This is the assessment error for the youngest at the school starting age discontinuity minus the assessment error for the oldest at the same $days_i = 0$. As this effect will be centered around zero, the resulting sign and magnitude of these deviations provides an unbiased estimate of the direction and magnitude of over- and under-assessment of ADHD symptoms by a child's age relative to their cohort. We further break this effect down by assessing school starting age effects by student gender and ADHD symptom sub-types. Sub-group analysis allows for differences in both the intercept and slope of the separate groups.

3.2.2 School and teacher mechanisms of assessment error

To investigate mechanisms that may serve to mitigate or exacerbate assessment errors, we compile school and class factors related to class size, teacher-rated classroom behaviour, the school-level proportion of students with physical, emotional and learning disorders and the proportion of students from low-income families. This last variable is defined as a household income below $\$30,000$. We also include a teacher's age, number of years of experience, gender and

[3]It is assumed that a child's true ADHD symptoms, similar to human capital traits, are observed by both teachers and parents with error. We assign parent-perceived human capital as: $h_i^P = h_i^* + \xi^P$. Teacher-perceived human capital is analogously $h_i^T = h_i^* + \xi^T$. Thus assessments are correlated and their measurement errors may also be correlated. Rearranging parent-perceived traits as $h_i^* = h_i^P - \xi^P$, we have true human capital dependent on perceived human capital, less the measurement error. We can then estimate teacher-perceived traits based on parent-perceived traits as:

$$h_i^T = [h_i^P - \xi^P] + \xi^T = h_i^P + [\xi^T - \xi^P]$$

The estimated residual from a ordinary least squares regression is then: $\eta_i = \hat{h}_i^T - \hat{\alpha} h_i^P$, which approximates to $\xi^T - \xi^P \sim N(0, \sigma^2)$.

special education training. The NLSCY asked teachers to specify their level of special education as none, a single class or course, a specialized degree, or master's degree. We first transform this variable into a binary indicator of having any special education training. We additionally include family income, mothers age and number of siblings to investigate family-level controls. We take average values of these variables across city, year and grade cohorts separately for children born one month before and also one month after their school entry eligibility cut-off. We then calculate the cohort school starting age effect as the difference between teacher assessments for those born one month before and one month after the cut-off date. To test how these factors impact assessment error, we estimate the following equation:

$$SSA_{ct} = \beta_0 + \beta_s School_{ct} + \beta_t Teacher_{ct} + \beta_f Family_{ct} + \beta_x X_{ct} + \lambda_t + \pi_c + \mu_{ct} \quad (4)$$

The calculated city-year-grade school starting age effect is contained in SSA_{ct}. Each β vector measures the impact of school, teacher and family variables contained within their respective matrices. Controls contained in a matrix X_{ct} relate to grade fixed effects, the proportion of females in a given cohort and indicators for the size and rurality of the city in question. Year and city-level fixed effects are measured through λ_t and π_c, respectively. We cluster the error term μ_{ct} at the city-year level. We include city-level fixed effects as there is potential for some school-board specific policies that lead school and teacher factors to differ, or families to opt to live in these cities, that may also impact the differential between the oldest and youngest per grade.

In line with Schwandt and Wuppermann (2016), we cannot consider these results to be casual but only suggestive as to specific variables of interest. We investigate any variables of particular importance within a stratified RD framework, as detailed in the previous section. Doing this will allow for a casual investigation into mechanisms, holding that there is no sorting around school start date by the investigated factors. We simultaneously test for this through a standard non-outcomes test, as presented in our mechanisms results section.

4 Results

Table 1 breaks down key variables of interest across groupings of students above (older) or below (younger) the median age in their grade, city and year. As expected, there are robust differences in age at school entry between the subsets. There are slightly more younger females than males, based on males being more likely to be held back from entering school, as seen in our data and in alignment with previous findings (Bassok & Reardon, 2013). Parents rate children as having a ADHD behavioural index score of 4.49 (3.17) out of a possible range of 0 to 16 and we do see a small difference between average parent assessments of younger and older children in a grade. Teachers using the same scale rate students with less misbehavior, at an average of 3.50 (3.69), but with a strong difference in assessments between younger and older students in a grade. Differences in teacher assessments for anxiety and depression are also present, but not for parent assessments. Differences in diagnosis by relative age grouping are similarly not found. This is not

Table 1: Summary Statistics by School Starting Age for Kindergarten Sample

	Full Sample	Older	Younger	Difference
Teacher Assessments				
ADHD	3.50 (3.69)	3.25 (3.60)	3.78 (3.76)	-0.53***
Anxiety/Depression	2.15 (2.59)	2.07 (2.51)	2.23 (2.67)	-0.16**
Conduct Disorder	1.05 (2.07)	1.03 (2.06)	1.06 (2.09)	-0.02
Parent Assessments				
ADHD	4.49 (3.17)	4.42 (3.14)	4.57 (3.19)	-0.15*
Anxiety/Depression	2.13 (2.13)	2.17 (2.15)	2.08 (2.11)	0.08
Conduct Disorder	1.55 (1.87)	1.57 (1.90)	1.54 (1.83)	0.03
Diagnosed ADHD	0.01 (0.08)	0.01 (0.09)	0.01 (0.07)	0.00
Age at School Entry	5.17 (0.30)	5.37 (0.24)	4.94 (0.18)	0.43***
Female	0.49 (0.50)	0.48 (0.50)	0.51 (0.50)	-0.02*
Log Household Income	10.76 (0.66)	10.75 (0.66)	10.76 (0.66)	-0.01
Late Enroll	0.16 (0.37)	0.31 (0.46)	0.00 (0.00)	0.31***
Early Enroll	0.03 (0.18)	0.00 (0.00)	0.07 (0.25)	-0.07***
N	7 510	2 945	3 574	

Note: Based on a kindergarten sample of the NLSCY. Older and younger columns split the sample into those above the median age at school start and less than the median age, respectively. All assessments categories are factors of the Children's Behavioural Scale. ADHD and anxiety/depression scales range from 0 to 16, while the conduct disorder scale has a maximum score of 12. Standard errors are reported in parentheses. The last column reports the older-younger difference, with significance values from a t-test of these two means corresponding to: * p<0.05, **p<0.01, *** p<0.001.

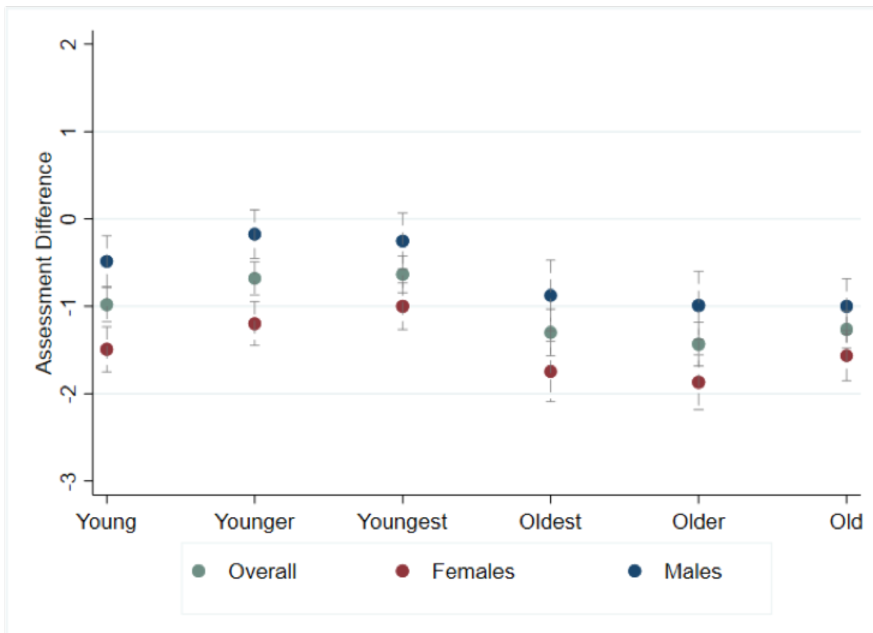
unexpected given that this kindergarten sample has less than 1% of students diagnosed with ADHD. [4]

In ascertaining the direction of ADHD assessment errors, we rely on teacher assessments, conditioned on parent assessments and by school starting age. Figure 1 gives a first look at the differences in average teacher assessments, less parent assessments by broad categories of school starting age. The youngest (oldest) students in a grade had birthdays falling within two months before (after) the school enrollment cut-off date. Younger (older) students were those with a birthday three to four months before (after) this same cut-off date, and young (old) students were those born five to six months before (after) the cut-off date. Parents, on average, rate their child as more misbehaved than teachers, as seen in Table 1. Therefore, differences between parent and teacher assessments across each age category

[4] Diagnosis of ADHD was only recorded for 2000-2002 in the NLSCY sample. A broader diagnosis category was used from years 1994-1998, asking if the child had an emotional, nervous, or psychological disorder, in which ADHD was categorized at the time. Previous work has found that the broad measure is consistent with known Canadian prevalence rates for ADHD at the time (Charach et al., 2010). We use this broader definition of emotional, nervous, or psychological disorder across all cycle years and combine this with Ritalin usage across all cycles and ADHD diagnosis in 2000-2002.

were negative. Teacher-reported scores were more closely aligned to parent-reported scores for the youngest students, as indicated by being closer to zero. The sizeable differences in parent-teacher assessments for the oldest suggests better ability to regulate classroom behaviour, based on relative maturity. Figure 1 also shows a clear gap in teacher-parent assessments from youngest to oldest students in a grade. Overall, younger students have a teacher-reported ADHD assessment 16% higher than their older peers, with a much smaller difference in parental assessments (Table 1). This gap increases to 20% when we narrow the comparator groups to include only those children born within two months of their school board-assigned eligibility date.

Figure 1: Differences in ADHD Assessment: Parent-Teacher Difference



Note: Each point is the average teacher minus parent assessment across gender and school starting age groups. The youngest (oldest) students had birthdays falling 2 months before (after) the enrollment cut-off date. Younger (older) were those with a birthday 3-4 months before (after) and young (old) for those 5-6 months before (after). Parents rate their child as more misbehaved than teachers, rendering negative differences across all groups.

Importantly, the discord between parent and teacher assessments of females is larger, regardless of school starting age. Teachers assess older females with 26% fewer symptoms than their younger counterparts, and they rate the oldest girls within a grade as having 48% fewer symptoms than the youngest girls. Without similar differences in parent assessments for females by school starting age, the drop in assessed ADHD severity between youngest and oldest females in a grade appears driven by teacher assessments. While there is a difference in the average assessed ADHD for males by school starting age group as well, it is comparatively smaller. Figure 1 gives credence to the hypothesis that teachers underweight female misbehaviour, especially in older females. This under-weighting is potentially due to females and older students' ability to self-regulate in school, but may also include some stereotyping components. Differences in assessment between these oldest versus youngest for grade students may also be based on other underlying characteristics of older versus younger students. Discerning which part of the teacher-parent difference is due to observed or perceived behaviour at school versus home remains the crux of this issue. We move to a casual identification strategy to answer this question.

4.1 School entry age effects in parent and teacher assessments

To understand assessment errors, we first confirm a previous hypothesis that teacher, but not parent assessments of kindergarten-aged students drive the school starting age effect. We present these results in Panel A of Table 2 over several specifications, including both global polynomial and local linear models. We further include key control variables and regional fixed effects to account for potential issues of monotonicity in assignment to treatment. Our results demonstrate a robust effect of school starting age on teacher-assessed ADHD in kindergarten across all specifications. The global polynomial estimate in column (1), incorporating all possible day_i observations from -180 to 180, has an estimated treatment effect of 1.37 (0.61). Global polynomial estimates are shown graphically in Appendix Figure A1. We however take local linear estimates to be better measures of the effect of interest (Gelman & Imbens, 2019). This effect increases under a local linear specification using observations within an optimal bandwidth, calculated using a two-sided mean squared error bandwidth estimator, of 30 days before the cut-off and 50 days after (Calonico, Cattaneo, Farrell, & Titiunik, 2018). In our fully specified model, including regional fixed effects (4), teachers assess the youngest students within a grade as having 1.90 (0.65) more symptoms than their oldest peers, half a standard deviation. Controlling for regional heterogeneity serves to increase the starting age effect, attenuating concerns that failure of the monotonicity assumption would unduly inflate this LATE. Controlling for parental assessments of ADHD does attenuate this effect somewhat in column (5). We thus take the final column results, a treatment effect of 1.83 (0.68), to be the best estimate of the LATE of school starting age on teacher ADHD assessments, holding home behaviour fixed. We see a near-zero impact of school starting age on parent-assessment of ADHD across all specifications and similarly no effect on ADHD diagnosis or treatment. A null effect on diagnosis is not unexpected in this kindergarten sample given the low diagnostic prevalence of ADHD.

Our result answer the Elder (2010) hypothesis that, when assessing behaviour in line with common diagnostic tools, teachers and not parents drive the school starting age effect. However, these results do not point to whether younger students within a grade are falsely over-assessed with ADHD-like symptoms or if older students within a grade are under-assessed. We turn to this next.

4.2 Over and under assessment in school starting age effects

To address direction of misidentification within the observed school starting age effects, we first predict teacher-assessed ADHD severity, using parent-assessed ADHD severity at home. We use a quadratic prediction of teacher assessments.[5] The estimated error from the prediction of teacher assessments based on parent assessments has a mean of zero and a standard deviation of 3.8 (Appendix Table A2).

In re-estimating our school entry age model using these estimated residuals as the outcome variable of interest, we

[5]Estimating equation: $h_i^T = \alpha_0 + \alpha_1 h_{ik}^P + \alpha_2 (h_{ik}^P)^2 + \eta_i$. We test other specifications for this equation, including a simple linear predictor and a fully non-linear factored index. See Appendix Table A2 for comparisons across three different prediction models. Based on several goodness-of-fit testing procedures, a quadratic model was determined to be the best approach.

Table 2: School Starting Age Effects on ADHD Assessments

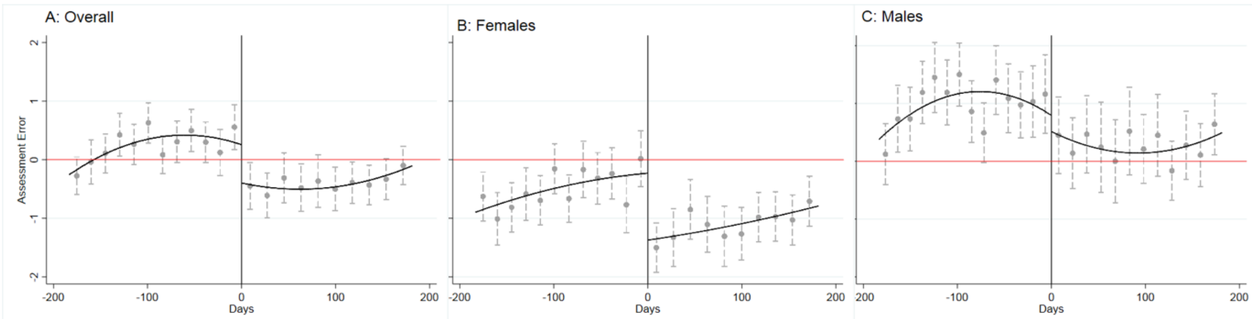
	Mean	(1)	(2)	(3)	(4)	(5)
	Panel A: Baseline Results					
Teacher Assessment	3.50	1.37*	1.65**	1.67**	1.90**	1.83**
	(3.69)	(0.61)	(0.61)	(0.65)	(0.65)	(0.68)
Parent Assessment	4.49	-0.29	-0.14	-0.04	0.15	-
	(3.17)	(0.55)	(0.50)	(0.52)	(0.53)	
Diagnosed/Treated	0.01	0.01	0.02	0.02	0.02	0.02
	(0.08)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Ritalin	0.004	0.000	0.005	0.005	0.006	0.005
	(0.07)	(0.01)	(0.004)	(0.004)	(0.004)	(0.004)
	Panel B: Residual Teacher Assessments					
Overall	0.00	1.55*	1.75**	1.65*	1.88**	-
	(3.44)	(0.65)	(0.66)	(0.71)	(0.70)	
Females (n = 3,701)	-0.73	3.20**	3.44**	3.43**	3.39**	-
	(2.99)	(1.00)	(1.05)	(1.09)	(1.10)	
Males (n = 3,809)	0.70	0.18	-0.17	-0.40	-0.53	-
	(3.69)	(1.20)	(1.05)	(1.15)	(1.10)	
Global Polynomial		x				
Local Linear			x	x	x	x
Controls				x	x	x
Regional Fixed Effects					x	x
Parent Assessed Score						x
N	7510	7510	7510	7510	7510	7510

Note: All reported estimates are bias-corrected local average treatment effects that first instrument for age at school entry in kindergarten. Local linear models estimated using an asymmetric two-sided mean squared error bandwidth selector. Controls include gender, mother age, number of siblings, log household income, and a control for junior kindergarten attendance. Significance level indicated by * p<0.05, **p<0.01, *** p<0.001.

expect a similar discontinuity to the one found in overall teacher assessments in Panel A. Panel B of Table 2 shows this assumption holds, with overall results almost identical to those in Panel A. From Column (4), the estimated assessment error is 1.88 (0.70), only 0.02 less than the effect found in overall teacher assessments. This decrease is the result of controlling for parent assessments in our quadratic prediction model. Across all specifications the estimated LATE increases or remains the same. A more robust picture of behavioural assessments emerges when separating the school starting age effect in assessment error by gender. Assessments of females drive the effect in full. The LATE estimated for females in Column (4) is 3.39 (1.10). The comparable local-linear school starting age effect for males is negative and not precisely estimated, suggesting a non-linear relationship.

Using this residual approach has the benefit of centering the school entry age effect around a mean of zero, from which we can break down the treatment effect into components of positive and negative deviations from zero in assessment error. If it holds that these deviations in assessment error are based solely on school entry age, we can interpret these as over and under assessment. Figure 2 demonstrates this direction within a global polynomial model. Overall, we see a positive assessment error for the youngest students in a grade, while their oldest day_i bin counterparts have a negative assessment error. Mirroring results found in Table 1, we see a substantial jump in ADHD

Figure 2: Teacher ADHD Assessment Error



Note: Each point shows estimation error binned by $days_i$, using an evenly spaced integrated mean squared error method, scaled by a factor of 2 to avoid over-smoothing (Calonico et al., 2018). Fitted lines represent use a quadratic polynomial. Error bars correspond to 95% confidence intervals, with standard errors clustered on $days_i$.

assessment error between the youngest and oldest females in a grade. Both older and younger females within a grade have negative assessment errors, outside of the very youngest females. At the same time, the assessment error for males, especially the youngest, is almost always positive. There is however limited evidence of a school starting age effect for kindergarten males in either the structural form (Figure 2c) or the fuzzy RD (Table 2). The school starting age effect for males is likely attenuated by a higher rate of red-shirting for males in kindergarten that may widen the variability in assessments for these young boys. The down-tick in average assessment error for males as the running variable approaches 0 from the left in Figure 2c demonstrates this.

To test the impact of red-shirting behaviour on the estimated school starting age effect and to estimate assessment errors on either side of the school starting age discontinuity, we re-run our local-linear model on a fully compliant sample of students. This sample excludes any students who enrolled either early or late. This sharp RD provides an average treatment effect (ATE) of age at school entry, which we expect to be smaller than the LATE, but more precisely estimated. This approach will also allow us to decompose the ATE into over and under-assessment by school starting age. Table 3 first provides the school starting age ATE in Panel A. The ATE is indeed smaller than the estimated LATE overall, but is more precisely estimated at 1.31 (0.33) versus 1.83 (0.68). We see this similarly for females, while ATE for males aligns with results found in the global polynomial model, with an effect size of 0.73 (0.59).

Point estimates of the assessment errors at the school starting age cut-point are provided in Panel B. These represent the treatment effects for having a birthday fall just before the cut-off date ($z \leq 0$), or just after ($z > 0$). The treatment effects on either side of the cut-point, τ^Y and τ^O result from being the youngest or oldest student in one's grade, respectively. These are used to construct $ATE = \tau^Y - \tau^O$. Overall for combined symptom-type ADHD, we find a rate of over-assessment for the youngest children in a grade of 0.51 (0.22) symptoms and an under-assessment of 0.80 (0.25) for the oldest kids. We see that the jump in assessment error between the youngest and oldest females in a grade still dominates the estimated effect, with an under-assessment of 1.94 (0.20) symptoms. The youngest females have a near-zero assessment error in the compliant sample at 0.02 (0.34). Over-assessment of the youngest males is significant, with 1.04 (0.33) additionally assigned symptoms based on relative immaturity.

Table 3: School Starting Age Effect and Intercept Point Estimates

	Panel A: Average Treatment Effect		
Sup-Group	Overall	Females	Males
ADHD Combined	1.31*** (0.33)	1.95*** (0.40)	0.73 (0.59)
Hyperactive	0.58** (0.19)	0.95** (0.32)	0.12 (0.38)
Inattentive	0.68** (0.21)	1.00*** (0.28)	0.30 (0.37)
Anxiety/Depression	0.57* (0.26)	1.84*** (0.32)	-0.95* (0.38)
Conduct Disorder	0.84*** (0.22)	0.45 (0.24)	1.19** (0.41)

	Panel B: Intercept Point Estimates					
	Overall		Females		Males	
Intercept Point	τ^Y	τ^O	τ^Y	τ^O	τ^Y	τ^O
ADHD Combined	0.51* (0.22)	-0.80** (0.25)	0.02 (0.34)	-1.94*** (0.20)	1.04** (0.33)	0.31 (0.49)
Hyperactive	0.12 (0.10)	-0.37*** (0.12)	-0.10 (0.25)	-1.04*** (0.11)	0.50** (0.20)	0.29 (0.25)
Inattentive	0.32** (0.12)	-0.33** (0.13)	0.10 (0.17)	-0.87*** (0.16)	0.56** (0.22)	0.13 (0.21)
Anxiety/Depression	0.18 (0.16)	-0.29+ (0.23)	0.67*** (0.24)	-0.97** (0.40)	-0.35* (0.16)	0.40 (0.45)
Conduct Disorder	0.64 (0.20)	-0.05 (0.17)	-0.04 (0.20)	-0.51*** (0.17)	1.36*** (0.37)	0.40+ (0.23)
N	7,109		3,516		3,593	

Note: Panel A reports estimated treatment effects of school starting age from a sharp regression discontinuity in a fully compliant kindergarten sample. Each reported coefficient is from a single regression based on disorder for the full sample or by gender subset. All coefficients are bias-corrected local-linear estimates using triangular kernel weighting. Panel B reports the paired intercept points from each of the regressions presented in Panel A. These are the intercept points for those born just before (τ_1) and just after (τ_0) the cut-point. Standard errors are reported in parentheses and are heteroskedasticity-robust and clustered on the day-level. Significance level presented as: + p<0.10, * p<0.05, **p<0.01, *** p<0.001.

Combining the intercept points in Panel B with the treatment effect in Panel A, we can calculate that over-assessment of the youngest students makes up 39% of the school starting age effect, while under-assessment of older students constitutes just over 61%. Furthermore, teachers appear to over-assess the youngest males in a grade, while simultaneously under-assessing symptoms in the oldest females. While robust measurement errors are present for both the youngest males and oldest females in a grade, the assessment error for these females is 1.87 times that of males.

If under assessments of females lead to fewer referrals, these results suggest a reason as to why ADHD is thought to be under-diagnosed and under-treated for females (Sciutto & Eisenberg, 2007; Swanson et al., 2013). These results are more pronounced when using a fuzzy RD on the full kindergarten sample to calculate the LATE (Appendix Table A3). The two-stage least squares design of the fuzzy RD limits the interpretation of intercept results, however.

Finding that under-assessment plays a larger role in ADHD evaluations appears at first glance to run counter to previous work positing over-diagnosis as the main culprit in the ADHD school starting age effect (Schwandt & Wuppermann, 2016). Since childhood ADHD is associated with higher rates of accidental injury (Nigg, 2013), if under-diagnosis were occurring, injury rates should be higher for older students not receiving appropriate interventions. However, the impact of ADHD treatment on injury occurrence is mixed (Marcus, Wan, Zhang, & Olfson, 2008; van den Ban et al., 2014; Dalsgaard, Leckman, Mortensen, Nielsen, & Simonsen, 2015), and differences in injury rate with appropriate treatment may not be assured. We also find an active gender component to under-assessment, but injuries for females with ADHD are less likely (Lindemann et al., 2017).

Moreover, as an additional indicator of over-diagnosis, Schwandt and Wuppermann (2016) show a marked correlation between a given country's overall ADHD prevalence rate and the estimated school starting age effect for that country. Our results do not negate this result but, given the robust gendered effects, help bolster the prevalence-school starting age relationship. Male diagnoses dominate most ADHD prevalence rates; for instance, for every ADHD-diagnosed female in Canada, three males are diagnosed (Brault & Lacourse, 2012). We find evidence of over-assessment for males regardless of school starting age in this study. Thus the high correlation between prevalence and school starting age is a potential sign of over-diagnosis for males. Most previous studies do not assess or do not find significant gender effects for diagnosis (e.g. Elder, 2010; Evans et al., 2010; Schwandt & Wuppermann, 2016). This lack of effect is likely due to a focus on diagnosis rates, which contain a minority count of females. Our use of survey data, though not without limitations, provides critical pre-diagnosis information from a nationally representative sample that is not available in administrative data files. Use of the NLSCY's parent and teacher assessments demonstrate the mechanics of assessments decision making that lead some children to be referred for formal assessments and diagnosed, while others are not. The only other study to use pre-diagnosis assessments also find strong gender differences in inattention/hyperactivity assessments in Denmark (Dee & Sievertsen, 2018), despite limited diagnosis differences by school starting age overall (Dalsgaard et al., 2012). Our results are larger than those found in Dee and Sievertsen (2018), however, different scales are used in different country contexts.

To verify the robustness of these results, we apply common tests related to differing jump points for the age cut-off and find no effect on results (results available upon request). We also re-run specifications excluding provinces in years in which registration cut-off dates were imposed at the school board level and find a small downward effect on results (1.60, s.e. 0.69). When we drop any child already diagnosed from the results we see no real change in the overall effect (1.77, s.e. 0.66).

4.2.1 Internalizing versus Externalizing Symptoms

To test ADHD assessment errors along internally versus externally presenting symptom sub-type, we separate the inattentiveness/hyperactivity scales for both parents and teachers into its sub-components. Re-estimating by symptom sub-type effectively splits the school entry age effects in half. Returning to Table 3, we see an effect of 0.58 (0.19) in hyperactive symptoms and 0.68 (0.21) in inattentive symptoms (Table 3). We find the school starting age effect to be higher for inattentiveness by gender subsets as well. Looking to assessment errors in Panel B, the under-assessment of older females (τ_0) is larger for hyperactivity, while the comparable over-assessment of the youngest males (τ_1) is larger for inattentiveness. Graphical representations by symptom type and gender are shown in Appendix Figure A2.

We also apply the same methods to estimate the error in teacher assessments of anxiety, depression, and conduct disorder using the CBS. We find these results to again fall along externally presenting versus internally presenting lines, with conduct disorders more likely to be externally presenting and anxiety or depression more internal. Significant assessment error effects are found in anxiety/depression for females, with an estimated jump in symptoms of 1.84 (0.32) between the youngest and oldest females in a grade. In Panel B we can see that the age effect is split over both under-assessment of the oldest females and the youngest females being rated as having higher levels of anxiety. For conduct disorder, we find any overall effect to be driven by school starting age effects for males. This treatment effect of 1.19 (0.41) for males is driven by an assessment error of 1.36 (0.37) more conduct disorder symptoms for the youngest males. The assessment error is however positive regardless of age. Baseline treatment effects on parent and teacher assessments of anxiety/depression and conduct disorder, along with estimates for indirect aggression are found in Appendix Table A4.

4.3 Assessment error mechanisms

To understand school or teacher factors that might serve to drive or mitigate assessment errors, in Table 4, we first assess the role of school, teacher and family factors factors directly on the estimated school entry age effect. We take the difference in average teacher assessments for those born in the month before and those born in the month directly after the school board cut-off date for a given city, by year and grade. We also run non-outcomes tests on each individual mechanism to ensure that jumps in the non-outcomes variable itself do not drive the found school starting age effect.

Addressing class size first, as was seen in Schwandt and Wuppermann (2016), we expect to see a positive relationship between class size and assessment error. Nevertheless, while we do see some evidence of a positive relationship between class size and the size of the school starting age effect, this effect disappears upon the inclusion of city-level fixed effects. Out of the school level factors, we find that the affluence level of the student body plays the most influential role. The proportion of children in a child's school from low-income families serves to increase assessment error in the age of school entry effect. A one percentage point increase in the proportion of students from families with household incomes below \$30,000 increases the school starting age effect by 0.03, or 1.6% of our estimated effect. This effect is estimated controlling for own family income. This result is reversed and slightly smaller for the proportion of

Table 4: Mechanisms of the School Starting Age Effects

School Factors	(1)	(2)	(3)	Non-Outcome Test
Class Size	0.07	0.08	0.00	0.58
	(0.17)	(0.18)	(0.22	(1.03)
Peer Behaviour	0.11	0.18	-0.50	0.18
	(0.61)	(0.61)	(0.92)	(0.16)
Physical Disabilities	0.01	0.00	0.01	-2.19
	(0.03)	(0.03)	(0.04)	(1.43)
Emotional Disorders	0.02	0.02	0.05	-3.64
	(0.05)	(0.05)	(0.05)	(1.52)
Learning Disorders	-0.01	-0.01	-0.01	-3.42
	(0.03)	(0.03)	(0.05)	(-0.59)
Low Income	0.03*	0.04*	0.03+	-3.84
	(0.02	(0.02)	(0.02)	(-0.37)
High Income	-0.03*	-0.04*	-0.02	-0.60
	(0.02)	(0.02)	(0.02)	(-0.07)
Teacher Factors				
Age	1.13	1.29	1.74*	-0.27*
	(0.90)	(0.91)	(0.87)	(0.13)
Experience	-0.01	-0.01	-0.01	-12.41
	(0.01)	(0.01)	(0.01)	(17.89)
Special Ed Training	-1.70**	-1.81**	-2.01**	0.04
	(0.63)	(0.64)	(0.63)	(0.17)
Family Factors				
Household Income	-1.07	-1.24	0.53	0.15
	(0.90)	(1.01)	(1.16)	(0.14)
Mother Age	-0.09	-0.04	-0.12	-0.19
	(0.15)	(0.15)	(0.20)	(1.33)
N Kids in Home	-0.42	-0.42	-0.91	-0.26
	(0.61	(0.65)	(0.84)	(0.24)
Controls	x	x	x	x
Year FE	x	x	x	x
Urban size		x	x	x
City FE			x	
N	250	250	250	-

Note: Controls in models (1) to (3): Female, grade. In the non-outcomes test robustness check column, controls include regional and year fixed effects and report bias corrected coefficients from local linear estimates using triangular kernel weighting on a fully compliant sample. Standard errors are clustered on city*year: + p<0.10, * p<0.05, **p<0.01, *** p<0.001.

high-income families in a school. Increasing the proportion of students in a school with families about $80,000 per year decreases the assessment error by 0.02, or 1.1%. Knowing the links between family disadvantage and non-cognitive skills development (Jacob, 2002; Autor et al., 2016), the affluence level in a school may approximate the non-cognitive skills development level of peers. While we do not find a strong effect of peer behaviour on city-year level gaps in assessments by school starting age in Table 4, we do see that a teacher's assessment of the individual student shifts by the reported level of peer misbehaviour (Appendix Figure A3). If a child is reported to be in a poorly behaved classroom, the overall average assessments of misbehaviour shift upwards for the individual child. Taken together, being in a school with higher levels of disadvantage may then generate an assessment error for the individual student. Whether this peer behaviour is an actual effect on the individual student's behaviour or perceived misbehaviour by the teacher remains unclear.

Adding to this, less affluent families are more likely to enroll their children in school as soon as they are eligible, and potentially less kindergarten ready (Dhuey et al., in press). More affluent families are conversely more likely to redshirt their children. In our sample, we see that the average income of families who redshirt their child is slightly higher, at around $4,000 more per year. Having more lower-income families in a school would thus lead to a lower incidence of red-shirting and more younger students within a grade. This strategy could serve to increase the school starting age effect in schools with a less affluent student body.

Table 5: School Starting Age Effect by Resource Mechanism

	Teacher Age	Special Education	Low income
SSA	1.87** (0.70)	1.88** (0.70)	1.86* (0.74)

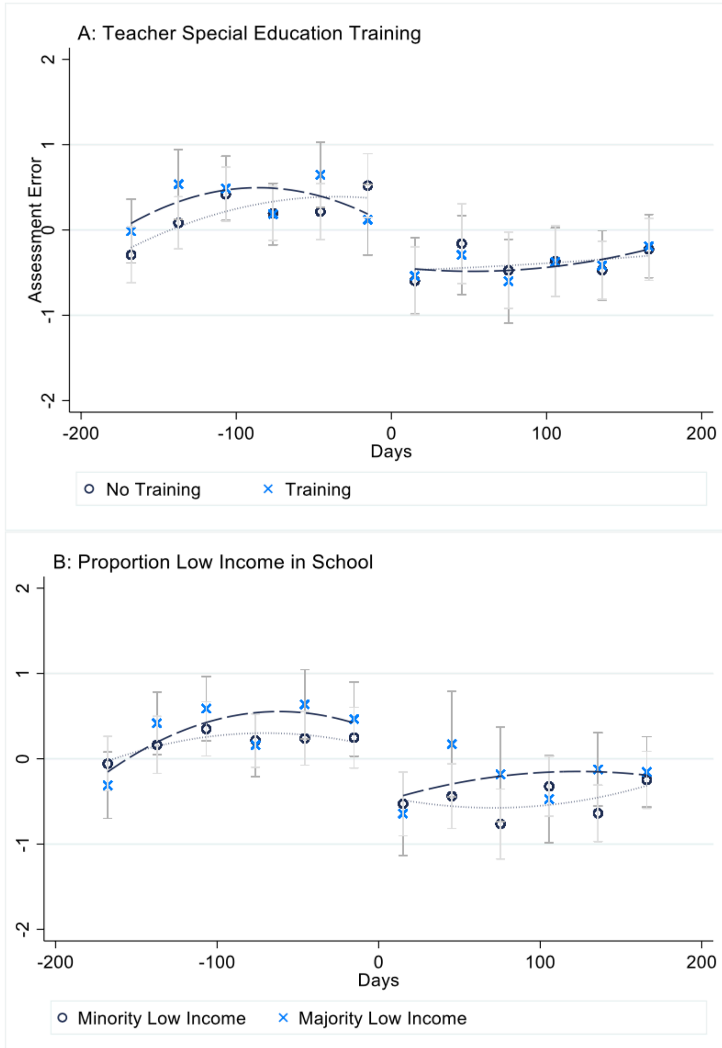
	Teacher Age				Special Education			Low income	
Subset	20-29	30-39	40-49	50+	No	Yes	Yes*	No	Yes
	Panel A: Full Sample								
SSA	1.92 (2.22)	1.72 (1.31)	1.23 (0.83)	2.75* (1.22)	2.75*** (0.82)	0.31 (0.71)	0.17 (0.80)	1.64 (1.03)	2.42* (1.05)
N	924	1,782	2,975	1,762	4,117	3,393	3,234	4,484	3,026
	Panel B: Compliant Sample								
SSA	1.29 (2.12)	2.11* (0.90)	1.68** (0.60)	2.42** (0.86)	2.00*** (0.49)	0.50 (0.44)	0.33 (0.53)	1.30** (0.48)	1.70* (0.70)
N	731	1,418	2,435	1,402	3,262	2,779	2,660	3,709	2,332

Note: Each reported coefficients are bias corrected from a single local linear regression using triangular kernel weighting. Standard errors are heteroskedasticity-robust and clustered on day-level. Controls are female, log household income, year number of children in household, maternal age, regional fixed effects. Under the special education grouping, Yes* shows a robustness check testing impact of including only those teachers with special education class or course but not full specialized degree or masters degree. Standard errors in parentheses: * p<0.05, ** p<0.01, *** p<0.001.

When we break down our estimation into minority and majority low-income schools, we see that being in a majority low-income school risks a higher school starting age effect (Table 5). While the effect is still present for minority low-income schools, it is smaller and not precisely estimated. However, when we use only a compliant sample, the effect sizes shrink for both, but at a faster rate for majority low-income schools. In the full sample, the majority low-income schools have an assessment error that is nearly 1.5 times that of the higher-income schools. This effect size difference falls to 1.3 for the compliant sample. We graphically compare assessment error differences across majority and minority low-income schools in Figure 3B. Figure 3B shows that the assessment error differences by school affluence are concentrated around the youngest students, again pointing towards the effect of higher red-shirting rates in more affluent schools as the cause of this income mechanism.

Returning to Table 4 and assessing teacher characteristics as mechanisms of assessment error, one factor, in particular, stands out: teacher special education training. Having special education training decreases the assessed

Figure 3: school starting age Effects by Mechanisms



Note: Each point is an average assessment error binned by $days_i$ before or after an eligibility cut-off date that a child's birthday falls, on a range of -180 to 180. Plots use an 1 month bins and a quadratic polynomial fit. Error bars represent 95% confidence intervals, with standard errors clustered on $days_i$.

school starting age effect by 2.01. As we estimate a school starting age effect of 1.83 (Table 2), this is the entirety of the effect. We return to our individual-level data and sub-set our sample based on whether the teacher had any special education training. Overall we see that including these key factors does not alter the estimated treatment effect (Table 5). In the sub-set sample, if a teacher has even some training, the school starting age effect disappears. This result holds in both the full sample and the compliant sample.

Graphically, we can see that the slope for ADHD assessments is less responsive to school starting age for the youngest students, suggesting these teachers are better able to condition their assessments on age (Figure 3A). This association is especially so for boys (Appendix Figure A4). It is, however, unclear if this special education effect makes a teacher better able to condition assessments on age or if teachers with special education training have some impact on the non-cognitive behaviours of their students directly(Jackson, 2012). Further to this, it is unclear if special education training itself impacts a teacher's ability to assess student behaviour. If there is selection into special education training based on unobservable factors, it may be that special education does not alter assessment ability, but that teachers less prone to assessment errors sought those designations.

As a robustness test, we remove any teachers who had a specialized degree or a masters degree in special education from the trained sample. We assume that there are less selective effects in a teacher opting to take a course or class over specializing in special education. Also, teachers with special education designations may be more likely to be assigned students with behavioural problems. When these teachers are removed, along with at least some selection effect, we see that the school starting age effect shrinks even further. Thus, while we cannot say definitively which direction this effect runs, these results give some basis for special education training impacting a teacher's ability to assess student behaviour or influence student behaviour for the youngest students.[6]

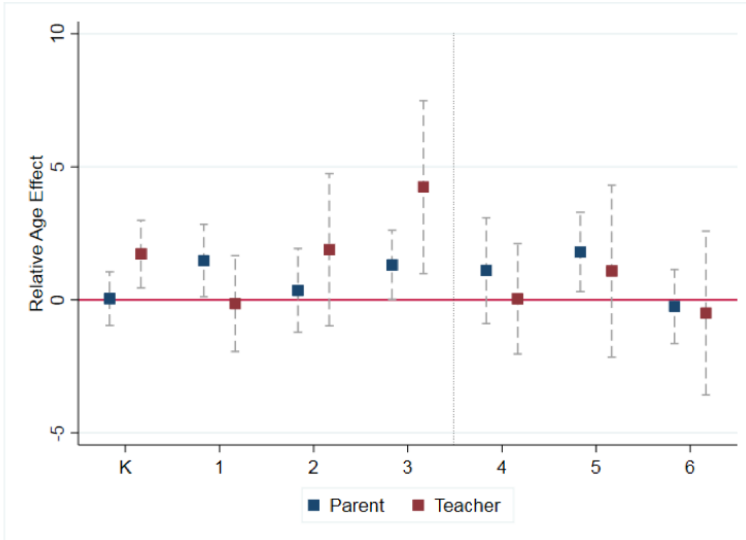
4.4 Aging out and Diagnosis

One final point for consideration is the duration of assessment errors based on school starting age. If teachers misconceive a student's relative maturity as evidence of ADHD, most likely through their ability to self-regulate between home and school environments, then age effects should dissipate as maturities converge with age. We extend our sample out to Grade 6 to disentangle this effect. Results from Figure 4 confirm that assessment errors based on school starting age in teacher assessments have a sharp drop off by Grade 4. Interestingly, we see that assessment errors in parent-reports emerge directly following kindergarten. Parent assessments continue to mirror teacher assessments, with a time lag, out until the Grade 4 drop off. This might indicate that teacher feedback regarding a child's behaviour causes parents to update their own beliefs, leading to a reinforcement of errors through both parent and teacher channels. Alternatively, it is plausible that a school starting age effect appears for parents after kindergarten because they have a broader reference group to compare their child to upon school enrollment. Nevertheless, the fact that the parent assessment so closely mirrors the teacher effect from the previous grade suggests a close relationship between the two.

When we break these longer-term effects down by gender and school starting age, we see that over-assessments of the youngest males rise after kindergarten but decline, starting in Grade 4 (Appendix Figure A5). For females, while the youngest females in a grade maintain near-zero assessment errors, the oldest females are consistently under-assessed until Grade 4. The Grade 4 drop-off in school starting age effects does suggest converging maturities. However, females do typically have later age of symptom onset (Quinn & Madhoo, 2014) which may influence under-assessment of females in this early grade sample. It is also of interest that ADHD and learning disorder testing and tracking into special education classes tends to begin in Grade 3. It is, however, difficult to test this pathway given that there are

[6]While we see a positive relationship between the age of the teacher and the school starting age effect, this may be driven by a non-continuous decrease in teacher age for the oldest students at the cut-point from the final column. This negative jump appears driven by a highly non-linear relationship upon visual inspection. We break down our model by teacher age categories and find that school starting age effects are larger for older teachers. We see some evidence of a negative impact of being a females teacher on the school starting age effects. However, given that over 98% of teachers in our kindergarten sample are females, we cannot discern precise results by gender matching.

Figure 4: Teacher and Parent school starting age Effects by Grade



Note: Each point is a local average treatment effect estimated using a local linear model and a two-sided MSE-optimal bandwidth selector. Error bars represent 95% confidence intervals, with standard errors clustered on $Days_i$.

no mandated times for testing initiation but only rules of thumb. We do see some evidence that school starting age effects in diagnosis begin to emerge after Grade 3, aligning with known patterns for testing across Canada (Appendix Table A5). The estimated local average treatment effect for diagnosis is a 4 percentage point difference in diagnoses. The average rate of diagnosis from Grade 3 to Grade 6 is 4 percent, indicating that all diagnoses come from the youngest in a grade. From our previous results, this would indicate that at least some over diagnosis is occurring in the youngest students. Since the diagnosed group is also mostly boys, any over-diagnosis is likely concentrated there. We do not see a similar differences in diagnosis rates by relative age for asthma, indicating that this effect is not likely driven by seasonal effects in diagnosis. Thus, while the teacher assessment errors may fade over time with student, given the instigation of a parent school starting age effect directly after kindergarten, these kindergarten assessment errors survive until the initiation of the common time of diagnosis three years later and directly impact who is diagnosed or missed.

5 Conclusion

The presented paper addresses the case of ADHD misidentification in childhood by assessing the degree of under and over assessment inherent within the well-known school starting age effect for pre-diagnosis assessments. Aligning with previously suggested mechanisms (i.e. Elder, 2010; Krabbe et al., 2014), we find that teacher assessments drive the school starting age effect in kindergarten. Contained within this assessment error, we show that over-assessment and under-assessment are both at play. Nevertheless, the role of under-assessment is more prominent. We present evidence that assessment errors based on school starting age are due predominantly to underassessment of the oldest students, particularly females, within a grade. We also show that females of all ages are under-assessed, outside of the youngest. At the same time, we can see that males in the classroom appear to be over-assessed by teachers. We argue that the school starting age effect beginning in kindergarten carries forward to subsequent years through a feedback loop with parents. This feedback loop instigates a selection effect in the referral process and, ultimately, diagnosis, with young-for-grade males the most likely to be diagnosed and potentially over-diagnosed.

These results firmly address a longstanding question in the school starting age literature on ADHD as to whether the diagnostic jump in ADHD, found in numerous countries, constitutes over or under-diagnosis. We posit both are present, but with under-assessment of the oldest students being a larger component. Whether the school starting age effect results solely from misattribution of relative maturity in both directions, or if stereotyping also plays a role remains to be answered substantively. The fact that under- or over-assessment appears to fall along gender-typical lines suggests the presence of some stereotyping. However, these gender-typical behaviours also fall along externalizing versus internalizing symptom presentation, which may also impact assessments based on symptom salience. Both of these factors aid in the potential over-diagnosis of males. In light of growing gendered differences in educational attainment, with males falling behind their female peers (Murnane, 2013), this paper offers the over-diagnosis of ADHD in males as a potential mechanism. Investigation of the precise role of ADHD on gendered educational differences remains a significant research gap.

We are also, to the best of our knowledge, the first study to test the impact of a comprehensive set of teacher and school characteristics as mechanisms of assessment errors. We additionally employ new strategies to test these critical factors at the individual level. The most substantial policy implication from our results suggests that teacher special education training attenuates assessment errors based on school starting age. With any form of training in special education, even a class or a course, the school starting age effect fell towards zero. As school starting age effects are found not just in ADHD but in learning disorder diagnosis, special education streaming, and overall educational outcomes, this finding has the potential for large scale welfare improvement. Exploring this teacher training mechanism further will provide fruitful grounds for future research.

Taking the findings presented in this paper, as ADHD affects 1 in 20 children worldwide (Faraone et al., 2003), the implications of misdiagnosis are considerable. With notable effects of ADHD on longer-term outcomes, appropriate diagnosis likely plays a vital role in long-term trajectories (Currie & Stabile, 2006; Fletcher, 2014). With previous work finding that younger children within a grade are also given more intensive medical treatment for ADHD (Layton et al., 2018) and more examinations and counselling (Balestra et al., 2017), addressing the causes of the school starting age assessment gap is also an issue of appropriate allocation of medical resources. The identification of under and over diagnosis along gender lines in this paper lays the groundwork for targeted interventions to improve the ADHD referral process and educational outcomes more generally. Finally, this paper offers teacher training as a potential policy intervention to curtail the significant reverberations of the school starting age effect in school.

References

Aizer, A. (2008). *Peer Effects, Institutions and Human Capital Accumulation: The Externalities of ADD.*

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (4th ed.).*

Autor, D., Figlio, D., Karbownik, K., Roth, J., & Wasserman, M. (2016, may). School quality and the gender gap in educational achievement. In *Am. econ. rev.* (Vol. 106, pp. 289–295). American Economic Association. doi: 10.1257/aer.p20161074

Autor, D., Figlio, D., Karbownik, K., Roth, J., & Wasserman, M. (2019, jul). Family disadvantage and the gender gap in behavioral and educational outcomes. *Am. Econ. J. Appl. Econ.*, *11*(3), 338–381. doi: 10.1257/app.20170571

Balestra, S., Eugster, B., & Liebert, H. (2017). *The Effect of School Starting Age on Special Needs Inci-dence and Child Development into Adolescence.* Retrieved from `www.RePEc.org`

Barua, R., & Lang, K. (2016). School entry, educational attainment, and quarter of birth: a cautionary tale of a local average treatment effect School Entry, Educational Attainment and Quarter of Birth: A Cautionary Tale of LATE. *Journal of Human Capital*, *10*(3), 347–376.

Bassok, D., & Reardon, S. F. (2013). "Academic Redshirting" in Kindergarten: Prevalence, Patterns, and Implications. *Educ. Eval. Policy Anal.*, *35*(3), 283–297. Retrieved from `https://journals.sagepub.com/doi/pdf/10.3102/0162373713482764`

Becker, G. S., Hubbard, W. H. J., & Murphy, K. M. (2010). The Market for College Graduates and the Worldwide Boom in Higher Education of Women. In *Am. econ. rev. pap. proc.* (Vol. 100, pp. 229–233). Retrieved from `http://www.aeaweb.org/articles.php?doi=10.1257/aer.100.2.229` doi: 10.1257/aer.100.2.229

Bharadwaj, P., Pai, M. M., & Suziedelyte, A. (2017, oct). Mental health stigma. *Econ. Lett.*, *159*, 57–60. doi: 10.1016/j.econlet.2017.06.028

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2011, may). Too young to leave the nest? The effects of school starting age. *Rev. Econ. Stat.*, *93*(2), 455–467.

Brault, M.-C., & Lacourse, É. (2012). Prevalence of Prescribed Attention-Deficit Hyperactivity Disorder Medications and Diagnosis Among Canadian Preschoolers and School-Age Children: 1994?2007. *Can. J. PsychiatryTheCJP.ca CanJPsychiatry*, *5757*(932), 93–101.

Brown, J., Quaskey, S., Rosenberg, L. A., Mellits, E. D., Denckla, M. B., Pierce, K., & Wolraich, M. L. (2001, oct). Clinical Practice Guideline: Treatment of the School-Aged Child With Attention-Deficit/Hyperactivity Disorder. *Pediatrics*, *108*(4), 1033–1044.

Bruchmüller, K., Margraf, J., & Schneider, S. (2012). Is ADHD Diagnosed in Accord With Diagnostic Criteria? Overdiagnosis and Influence of Client Gender on Diagnosis. *J. Consult. Clin. Psychol.*, *80*(1), 128–138.

Buckles, K. S., & Hungerman, D. M. (2013). Season of Birth and Later Outcomes: Old Questions, New Answers. *Rev. Econ. Stat.*, *95*(3), 711–724.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2018). RDROBUST: Stata module to provide robust data-driven inference in the regression-discontinuity design. *Statistical Software Components*(S458483).

Cascio, E. U., & Schanzenbach, D. W. (2016, jul). First in the class? Age and the education production function. *Educ. Financ. Policy*, *11*(3), 225–250.

Chan, E., Hopkins, M. R., Perrin, J. M., Herrerias, C., & Homer, C. J. (2005, jul). Diagnostic practices for attention deficit hyperactivity disorder: a national survey of primary care physicians. *Ambul. Pediatr.*, *5*(4), 201–8.

Charach, A., & Fernandez, R. (2013, jul). Enhancing ADHD Medication Adherence: Challenges and Opportunities. *Curr. Psychiatry Rep.*, *15*(7), 371.

Charach, A., Lin, E., & To, T. (2010, jun). Evaluating the Hyperactivity/Inattention Subscale of the National Longitudinal Survey of Children and Youth. *Heal. reports*, *21*(2), 43–50.

Chorniy, A., & Kitashima, L. (2016, dec). Sex, drugs, and ADHD: The effects of ADHD pharmacological treatment on teens' risky behaviors. *Labour Econ.*, *43*, 87–105. doi: 10.1016/J.LABECO.2016.06.014

Currie, J., & Schwandt, H. (2013). Within-mother analysis of seasonal patterns in health at birth. *Proceedings of the National Academy of Sciences*, *110*(30), 12265–12270.

Currie, J., & Stabile, M. (2006). Child mental health and human capital accumulation: The case of ADHD. *J. Health Econ.*, *25*(6), 1094–1118.

Currie, J., Stabile, M., & Jones, L. (2014). Do stimulant medications improve educational and behavioral outcomes for children with ADHD? *J. Health Econ.*, *37*, 58–69. doi: 10.1016/j.jhealeco.2014.05.002

Daley, D., & Birchwood, J. (2010, jul). ADHD and academic performance: why does ADHD impact on academic performance and what can be done to support ADHD children in the classroom? *Child. Care. Health Dev.*, *36*(4), 455–464.

Dalsgaard, S., Humlum, M. K., Nielsen, H. S., & Simonsen, M. (2012). Relative standards in ADHD diagnoses: The role of specialist behavior. *Econ. Lett.*, *117*(3), 663–665.

Dalsgaard, S., Leckman, J., Mortensen, P., Nielsen, H. S., & Simonsen, M. (2015, aug). Effect of drugs on the risk of injuries in children with attention deficit hyperactivity disorder: a prospective cohort study. *The Lancet Psychiatry*, *2*(8), 702–709. doi: 10.1016/S2215-0366(15)00271-0

Dee, T. S., & Sievertsen, H. H. (2018, may). The gift of time? School starting age and mental health. *Heal. Econ. (United Kingdom)*, *27*(5), 781–802.

Dhuey, E., Figlio, D., Karbownik, K., & Roth, J. (in press). School starting age and cognitive development. *Journal of Policy Analysis and Management*, *0*(0). doi: 10.1002/pam.22135

Dhuey, E., & Lipscomb, S. (2010). Disabled or Young? Relative Age and Special Education Diagnoses in Schools. *Econ. Educ. Rev.*, *29*(5), 857–872.

Dickert-Conlin, S., & Elder, T. (2010, oct). Suburban legend: School cutoff dates and the timing of births. *Econ. Educ. Rev.*, *29*(5), 826–841. doi: 10.1016/j.econedurev.2010.03.004

Doyle, O., Harmon, C. P., Heckman, J. J., & Tremblay, R. E. (2009). Investing in early human development: Timing and economic efficiency. *Econ. Hum. Biol.*, *7*(1), 1–6. doi: 10.1016/j.ehb.2009.01.002

Edmunds, A., & Martsch-Litt, S. (2008). Exceptionality Education International ADHD Assessment and Diagnosis in Canada: An Inconsistent but Fixable Process. *Except. Educ. Int.*, *18*(2), 2–23.

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

Elder, T. E., & Lubotsky, D. H. (2009). Kindergarten Entrance Age and Children's Achievement Impacts of State Policies, Family Background, and Peers. *J. Hum. Resour.*, *44*(3), 642–683.

Evans, W. N., Morrill, M. S., & Parente, S. T. (2010). Measuring Inappropriate Medical Diagnosis and Treatment in Survey Data: The Case of ADHD among School-Age Children. *J. Health Econ.*, *29*(5), 657–673.

Faraone, S. V., Sergeant, J., Gillberg, C., & Biederman, J. (2003, jun). The worldwide prevalence of ADHD: is it an American condition? *World Psychiatry*, *2*(2), 104–13.

Feng, L., & Sass, T. R. (2013, oct). What makes special-education teachers special? Teacher training and achievement of students with disabilities. *Econ. Educ. Rev.*, *36*, 122–134. doi: 10.1016/j.econedurev.2013.06.006

Figlio, D. (2007, oct). Boys Named Sue: Disruptive Children and Their Peers. *Educ. Financ. Policy*, *2*(4), 376–394.

Fletcher, J. (2010). Spillover Effects of Inclusion of Classmates with Emotional Problems. *J. Policy Anal. Manag.*, *29*(1), 69–83.

Fletcher, J. (2014). The effects of childhood ADHD on adult labor market outcomes. *Health Econ.*, *23*(2), 159–181.

Gaub, M., & Carlson, C. L. (1997, aug). Gender Differences in ADHD: A Meta-Analysis and Critical Review. *J. Am. Acad. Child Adolesc. Psychiatry*, *36*(8), 1036–1045.

Gelman, A., & Imbens, G. (2019, jul). Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs. *J. Bus. Econ. Stat.*, *37*(3), 447–456. doi: 10.1080/07350015.2017.1366909

Gershon, J. (2002, jan). A Meta-Analytic Review of Gender Differences in ADHD. *J. Atten. Disord.*, *5*(3), 143–154.

Hahn, J., Todd, P., & Klaauw, W. (2001, jan). Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica*, *69*(1), 201–209. Retrieved from `http://www.ifpri.org/publication/progresa-and-its-impacts-welfare-rural-households-mexico` doi: 10.1111/1468-0262.00183

Hanushek, E. A. (2011, jun). The economic value of higher teacher quality. *Econ. Educ. Rev.*, *30*(3), 466–479. doi: 10.1016/j.econedurev.2010.12.006

Heckman, J. (2007). The economics, technology, and neuroscience of human capability formation. *Natl. Acad. Sci. USA*, *104*(33), 13250–13255.

Hinshaw, S., & Scheffler, R. M. (2014). *The ADHD explosion: Myths, medication, money, and today's push for performance.* New York, NY: US: Oxford University Press.

Hinshaw, S., Scheffler, R. M., Fulton, B., Aase, H., Banaschewski, T., Cheng, W., ... Weiss, M. D. (2011). International Variation in Treatment Procedures for ADHD: Social Context and Recent Trends. *Psychiatr. Serv.*, *62*(4), 45–59.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *J. Econom.*, *142*, 615–635.

Jackson, C. K. (2012). *NON-COGNITIVE ABILITY, TEST SCORES, AND TEACHER QUALITY: EVIDENCE FROM 9TH GRADE TEACHERS IN NORTH CAROLINA.* Retrieved from http://www.nber.org/papers/w18624

Jacob, B. A. (2002, dec). Where the boys aren't: Non-cognitive skills, returns to school and the gender gap in higher education. *Econ. Educ. Rev.*, *21*(6), 589–598. doi: 10.1016/S0272-7757(01)00051-6

Jensen, P. S., Martin, D., & Cantwell, D. P. (1997). Comorbidity in adhd: implications for research, practice, and dsm-v. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(8), 1065–1079.

Kahneman, D., & Tversky, A. (1972). Subjective Probability: A Judgment of Representativeness. *Cogn. Psychol.*, *3*(430454).

Krabbe, E., Thoutenhoofd, E., Conradi, M., Pijl, S., & Batstra, L. (2014, oct). Birth month as predictor of ADHD medication use in Dutch school classes. *Eur. J. Spec. Needs Educ.*, *29*(4), 571–578.

Langberg, J. M., Froehlich, T. E., Loren, R. E., Martin, J. E., & Epstein, J. N. (2008, apr). Assessing children with ADHD in primary care settings. *Expert Rev. Neurother.*, *8*(4), 627–641.

Layton, T. J., Barnett, M. L., Hicks, T. R., & Jena, A. B. (2018, nov). Attention DeficitHyperactivity Disorder and Month of School Enrollment. *N. Engl. J. Med.*, *379*(22), 2122–2130.

Lindemann, C., Langner, I., Banaschewski, T., Garbe, E., Mikolajczyk, R. T., Garbe, E., ... Mikolajczyk, R. T. (2017, oct). The Risk of Hospitalizations with Injury Diagnoses in a Matched Cohort of Children and Adolescents with and without Attention Deficit/Hyperactivity Disorder in Germany: A Database Study. *Front. Pediatr.*, *5*. Retrieved from http://journal.frontiersin.org/article/10.3389/fped.2017.00220/full doi: 10.3389/fped.2017.00220

Lubotsky, D., & Kaestner, R. (2016). *Do 'Skills Beget Skills'? Evidence on the Effect of Kindergarten Entrance Age on the Evolution of Cognitive and Non-cognitive Skill Gaps in Childhood.*

Marcus, S. C., Wan, G. J., Zhang, H. F., & Olfson, M. (2008, jul). Injury among stimulant-treated youth with ADHD. *J. Atten. Disord.*, *12*(1), 64–9. doi: 10.1177/1087054707305168

Morrow, R. L., Garland, E. J., Wright, J. M., Maclure, M., Taylor, S., & Dormuth, C. R. (2012, apr). Influence of relative age on diagnosis and treatment of attention-deficit/hyperactivity disorder in children. *CMAJ*, *184*(7), 755–62.

Moses, T. (2010, apr). Being treated differently: Stigma experiences with family, peers, and school staff among adolescents with mental health disorders. *Soc. Sci. Med.*, *70*(7), 985–993.

Murnane, R. J. (2013, jun). U.S. high school graduation rates: Patterns and explanations. *J. Econ. Lit.*, *51*(2), 370–422. doi: 10.1257/jel.51.2.370

NCES. (2018, may). *The Condition of Education 2018* (Tech. Rep.). Retrieved from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2018144

Nigg, J. T. (2013, mar). Attention-deficit/hyperactivity disorder and adverse health outcomes. *Clin. Psychol. Rev.*, *33*(2), 215–28. doi: 10.1016/j.cpr.2012.11.005

Ohan, J. L., Visser, T. A., Strain, M. C., & Allen, L. (2011, feb). Teachers' and education students' perceptions of and reactions to children with and without the diagnostic label ADHD. *J. Sch. Psychol.*, *49*(1), 81–105.

Ohan, J. L., & Visser, T. A. W. (2009, aug). Why Is There a Gender Gap in Children Presenting for Attention Deficit/Hyperactivity Disorder Services? *J. Clin. Child Adolesc. Psychol.*, *38*(5), 650–660.

Pelham, W. E., Gnagy, E. M., Greiner, A. R., Hoza, B., Hinshaw, S., Swanson, J. M., ... McBurnett, K. (2000, dec). Behavioral versus behavioral and pharmacological treatment in ADHD children attending a summer treatment program. *J. Abnorm. Child Psychol.*, *28*(6), 507–525. doi: 10.1023/A:1005127030251

Polanczyk, G. V., Willcutt, E. G., Salum, G. A., Kieling, C., & Rohde, L. A. (2014). Adhd prevalence estimates across three decades: an updated systematic review and meta-regression analysis. *International journal of epidemiology*, *43*(2), 434–442.

Quinn, P. O., & Madhoo, M. (2014). A review of attention-deficit/hyperactivity disorder in women and girls: uncovering this hidden diagnosis. *Prim. care companion CNS Disord.*, *16*(3).

Sax, L., & Kautz, K. J. (2003). Who first suggests the diagnosis of attention-deficit/hyperactivity disorder? *Ann. Fam. Med.*, *1*(3), 171–4.

Schnoes, C., Reid, R., Wagner, M., & Marder, C. (2006). ADHD Among Students Receiving Special Education Services: A National Survey. *Except. Child.*, *72*(4), 483–496.

Schwandt, H., & Wuppermann, A. (2016). The youngest get the pill: ADHD misdiagnosis in Germany, its regional correlates and international comparison. *Labour Econ.*, *43*, 72–86.

Sciutto, M. J., & Eisenberg, M. (2007). Evaluating the Evidence For and Against the Overdiagnosis of ADHD. *J. Atten. Disord.*, *11*(2), 106–113.

Statistics Canada. (1996). *National Longitudinal Survey of Children: Overview of Survey Instruments for 1994-1995 Data Collection Cycle 1* (Tech. Rep.). Ottawa: Ministry of Industry.

Statistics Canada. (1997). *National Longitudinal Survey of Children: User's Handbook and Microdata Guide Cycle 1* (Tech. Rep.). Ottawa: Ministry of Industry.

Statistics Canada. (2002). *Microdata User Guide National Longitudinal Survey of Children and Youth 1998 to 1999* (Tech. Rep.). Ottawa: Ministry of Industry.

Swanson, J. M., Lakes, K. D., Wigal, T. L., & Volkow, N. D. (2013). *Multiple Origins of Sex Differences in Attention Deficit Hyperactivity Disorder.* Springer, Berlin, Heidelberg.

van den Ban, E., Souverein, P., Meijer, W., van Engeland, H., Swaab, H., Egberts, T., & Heerdink, E. (2014, feb). Association between ADHD drug use and injuries among children and adolescents. *Eur. Child Adolesc. Psychiatry*, *23*(2), 95–102. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/23733150` `http://link.springer.com/10.1007/s00787-013-0432-8` doi: 10.1007/s00787-013-0432-8

Visser, S. N., Danielson, M. L., Bitsko, R. H., Holbrook, J. R., Kogan, M. D., Ghandour, R. M., ... Blumberg, S. J. (2014). Trends in the Parent-Report of Health Care Provider-Diagnosed and Medicated Attention-Deficit/Hyperactivity Disorder: United States, 2003-2011. *J. Am. Acad. Child Adolesc. Psychiatry*, *53*(1), 34–46.e2.

Appendix

A DSM-IV Criteria for ADHD

Inattention Questions:

1. Often has trouble holding attention on tasks or play activities.
2. Often fails to give close attention to details or makes careless mistakes in schoolwork, at work, or with other activities.
3. Often does not seem to listen when spoken to directly.
4. Often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace (e.g., loses focus, side-tracked).
5. Often has trouble organizing tasks and activities.
6. Often avoids, dislikes, or is reluctant to do tasks that require mental effort over a long period of time (such as schoolwork or homework).
7. Often loses things necessary for tasks and activities (e.g. school materials, pencils, books, tools, wallets, keys, paperwork, eyeglasses, mobile telephones).
8. Is often easily distracted
9. Is often forgetful in daily activities.

Hyperactivity and Impulsivity Questions:

1. Often fidgets with or taps hands or feet, or squirms in seat.
2. Often leaves seat in situations when remaining seated is expected.
3. Often runs about or climbs in situations where it is not appropriate (adolescents or adults may be limited to feeling restless).
4. Often unable to play or take part in leisure activities quietly.
5. Is often "on the go" acting as if "driven by a motor".
6. Often talks excessively.
7. Often blurts out an answer before a question has been completed.
8. Often has trouble waiting his/her turn.
9. Often interrupts or intrudes on others (e.g., butts into conversations or games)

Each requires:

- Six or more symptoms for children up to age 16, or five or more for adolescents 17 and older and adults.
- Several inattentive or hyperactive-impulsive symptoms were present before age 12 years.
- Several symptoms are present in two or more setting.
- Clear evidence that the symptoms interfere with, or reduce the quality of, social, school, or work functioning.
- The symptoms are not better explained by another mental disorder.

NLSCY Parent and Teacher Reported Hyperactivity - Inattention Questions

1. How often would you say that [child/student] can't sit still, is restless or hyperactive?
2. How often would you say that [child/student] fidgets?
3. How often would you say that [child/student] is impulsive, acts without thinking?
4. How often would you say that [child/student] has difficulty awaiting turn in games or groups?
5. How often would you say that [child/student] cannot settle to anything for more than a few moments?
6. How often would you say that [child/student] is distractible, has trouble sticking to an activity?
7. How often would you say that [child/student] can?t concentrate, can?t pay attention for long?
8. How often would you say that [child/student] is inattentive?

Equating to a total score of 1-16 weighted by responses of: Never/Not True, Sometimes/Somewhat True or Often/Very True.

B Provincial Age at School Entry Cut-off Rules

Table A1: Canadian Provincial Eligibility Rules: 1994 to 2002

	1994	1995	1996	1997	1998	1999	2000	2001	2002
Alberta	SB	SB	SB	SB	SB	SB	SB	SB	SB
BC	Dec31	Dec31	Dec31	Dec31	Dec31	Dec31	Dec31	Dec31	Dec31
Manitoba	Dec1*	Dec1*	Dec31	Dec31	Dec31	Dec31	Dec31	Dec31	Dec31
New Brunswick	Dec31	Dec31	Dec31	Dec31	Dec31	Dec31	Dec31	Dec31	Dec31
Newfoundland	Dec31	Dec31	Dec31	Dec31	Dec31	Dec31	Dec31	Dec31	Dec31
Nova Scotia	Oct1	Oct1	Oct1	Oct1	Oct1	Oct1	Oct1	Oct1	Oct1
Ontario	Dec31	Dec31	Dec31	Dec31	Dec31	Dec31	Dec31	Dec31	Dec31
PEI	Jan31	Jan31	Jan31	Jan31	Jan31	Jan31	Jan31	Jan31	Jan31
Quebec	Sept30	Sept30	Sept30	Sept30	Sept30	Sept30	Sept30	Sept30	Sept30
Saskatchewan	SB	SB	SB	SB	SB	SB	SB	SB	SB

Note: SB = School board; SSD = School start date. *Functionally entry date begins on December 1 in these years but school board could set it to December 31. In Alberta, March 1 is the province-wide cut-off, though schools were allowed to set earlier dates, typically December 31. In 2018 the province-wide date moved to December 31. January 31 is province-wide date in Saskatchewan, with school board specific dates set earlier if desired.

C Appendix Tables

Table A2: Estimated Residuals in Teacher Assessments

	Linear Model		Quadratic Model		Factored Model	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
	Panel A: Kindergarten Sample					
Predicted Y	3.50	1.31	3.50	1.32	3.50	1.33
Residual	0.00	3.45	0.00	3.44	0.00	3.44
R-squared		0.13		0.13		0.13
BIC		39,911		39,901		40,012
AIC		39,897		39,880		39,894
N		7,510		7,510		7,510
	Panel B: Post-Kindergarten Sample					
Predicted Y	3.49	1.48	3.49	1.48	3.49	1.48
Residual	0.00	3.36	0.00	3.36	0.00	3.36
R-squared		0.17		0.17		0.17
BIC		106,960		106,903		107,023
AIC		106,944		106,880		106,889
N		20,315		20,315		20,315

Table A3: Direction of Assessment Error by School Starting Age in Full Sample

	Overall		Females		Males	
	Young	Old	Young	Old	Young	Old
Overall	1.88**		3.39**		-0.53	
	(0.70)		(1.10)		(1.10)	
	0.60	-1.72***	0.32	-5.32***	0.85	2.26***
	(0.42)	(0.45)	(0.70)	(0.64)	(0.60)	(0.72)
ADHD:HI	0.72		1.72**		-0.56	
	(0.38)		(0.67)		(0.53)	
	0.15	-0.78***	-0.02	-3.12***	0.37	1.67***
	(0.20)	(0.24)	(0.43)	(0.39)	(0.26)	(0.38)
ADHD:IN	1.00**		1.36**		-0.09	
	(0.37)		(0.42)		(0.61)	
	0.34	-0.80***	0.08	-2.21***	0.41	0.87*
	(0.23)	(0.22)	(0.26)	(0.26)	(0.37)	(0.37)
N	7,510		3,701		3,809	

Note: Table reports bias corrected coefficients from local linear estimates using triangular kernel weighting in a fuzzy regression discontinuity. Standard errors are heteroskedasticity-robust and clustered on day-level. Standard errors in parentheses: + p<0.10, * p<0.05, **p<0.01, *** p<0.001.

Table A4: School Starting Age Effects in Other Common Co-morbidities

Disorder	Anxiety	Conduct	Indirect Aggression
	Panel A: Overall Assessments		
Teacher	0.46(0.35)	1.15*(0.45)	0.22(0.35)
Parent	-0.23(0.34)	0.09(0.33)	-0.05(0.25)
Error	0.57*(0.26)	0.84*** (0.22)	0.07(0.18)
N	7,109	7,078	5,431
	Panel B: Females Only		
Teacher	2.46***(0.66)	0.89*(0.45)	0.51(0.63)
Parent	-0.55 (0.49)	0.52*(0.26)	0.06(0.32)
Error	1.84*** (0.32)	0.45 (0.24)	0.46 (0.35)
N	3,516	3,519	2,693
	Panel C: Males Only		
Teacher	-1.45*(0.66)	1.37*(0.66)	-0.27(0.47)
Parent	0.45(0.41)	-0.35(0.49)	-0.30(0.34)
Error	-0.95*(0.38)	1.19** (0.41)	-0.34 (0.27)
N	3,593	3,559	2,738

Note: Each reported coefficient is from a separate local-linear regression of assessment based on a school starting age. Results reported for teacher assessments, parent assessments, and residual teacher assessments conditioned on parent assessments. Results in Panel A are for the full sample, Panels B and C report female and male estimates separately. In each regression controls include log household income, year of assessment, number of children in household, maternal age, regional fixed effects. Gender controls included in overall regressions. All coefficients are bias-corrected from local linear estimates using triangular kernel weighting. Bandwidth selected using an MSE optimal two-sided bandwidth estimator. Standard errors in parentheses are heteroskedasticity-robust and clustered on day-level: * p<0.05, **p<0.01, *** p<0.001.

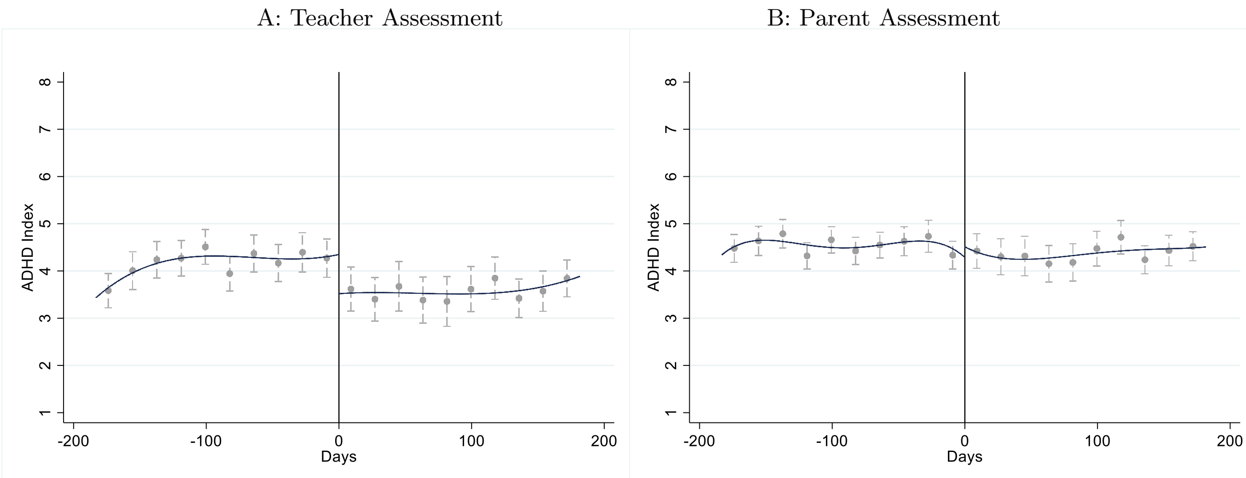
Table A5: School Starting Age Effect by Diagnosis

	ADHD		Asthma		
	Mean	LATE	Mean	LATE	N
	Full Sample				
Kindergarten	0.01 (0.08)	0.02 (0.02)	0.13 (0.34)	0.12 (0.07)	7,505
Post Grade 3	0.05 (0.21)	0.03 (0.03)	0.16 (0.37)	-0.01 (0.04)	15,578
	Compliant Sample				
Kindergarten	0.01 (0.08)	0.01 (0.01)	0.13 (0.34)	0.03 (0.04)	6,036
Post Grade 3	0.04 (0.20)	0.04* (0.02)	0.16 (0.37)	0.00 (0.03)	13,520

Note: Controls: female, log household income, year number of children in household, maternal age, regional fixed effects. Coefficients are bias corrected from local linear estimates using triangular kernel weighting. Bandwidth selected using an MSE optimal two-sided bandwidth estimator. Standard errors in parentheses are heteroskedasticity-robust and clustered on day-level: * p<0.05, **p<0.01, *** p<0.001.

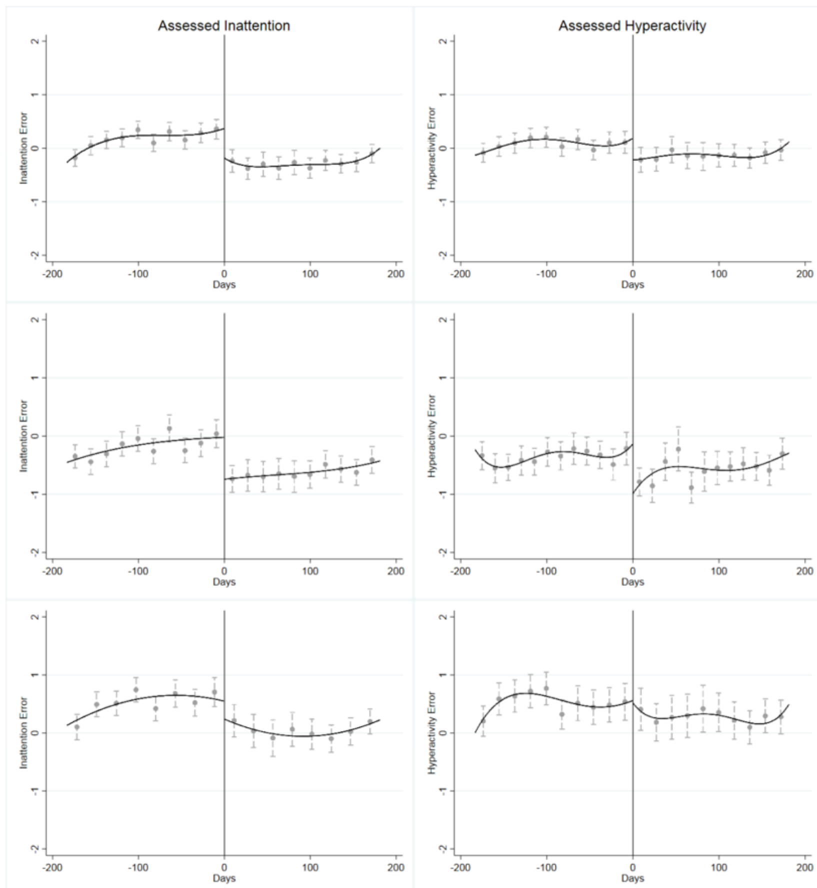
D Appendix Figures

Figure A1: Parent and Teacher Assessments of ADHD by school starting age



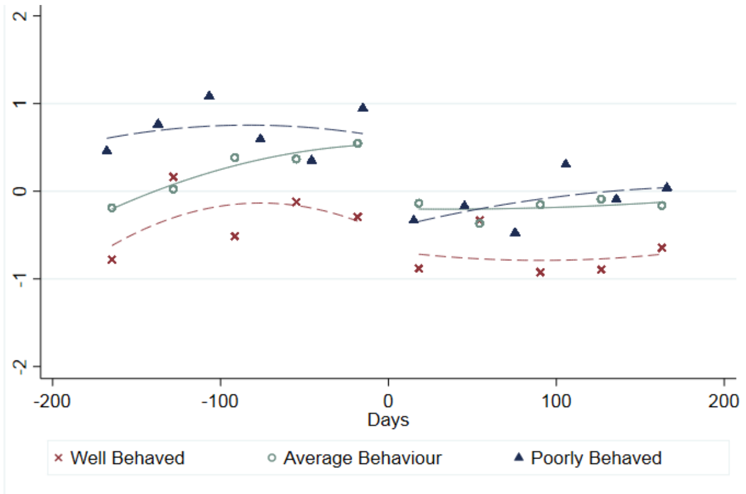
Note: The sample is based on all kindergarten-aged students in the NLSCY from years 1994 to 2002. Panel A represents teacher assessments of student ADHD severity using an ADHD behavioural index on a scale of 0 to 16. Panel B represents a parent assessments of the same child and using the same scale. The running variable $Days_i$ is the number of days before or after an eligibility cut-off date that a child's birthday falls, on a range of -180 to 180. Each point represents an averaged assessed ADHD score by binned $Days_i$. We use 10 day bins on both the right and left hand side in accordance with an integrated mean squared error (IMSE) evenly-spaced optimal bin selection method. We scale this by a factor of 2 to avoid over-smoothing. The fitted lines are flexible 3rd (Teacher) and 4th (Parent) degree global polynomial estimations fitted on either side of the eligibility cut-off. Error bars represent 95% confidence intervals, with standard errors clustered on $Days_i$. This graphical representation demonstrates school starting age effects for a fully compliant sample. fuzzy RD estimates from Table 2 demonstrate that these effects grow when accounting for non-random late or early enrollment.

Figure A2: Gender and Symptom Sub-Groups



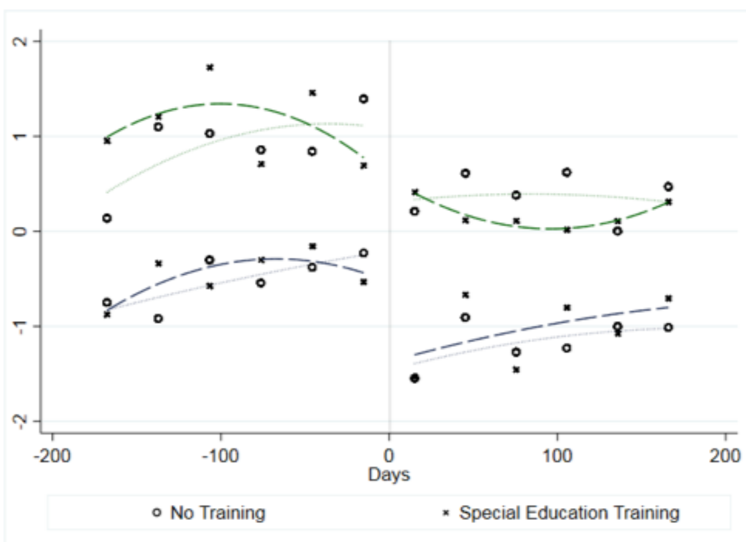
Note: school starting age effects calculated by symptom sub-type overall (Top Panel), for females (middle panel) and males (bottom). Panels show mean assessment errors binned by $Days_i$ before or after an eligibility cut-off date that a child's birthday falls, on a range of -180 to 180. Plots use an IMSE evenly-spaced optimal bin selection method to calculate bin size and flexible global polynomial fits. Error bars represent 95% confidence intervals, with standard errors clustered on $Days_i$.

Figure A3: Assessment Error by Peer Behaviour



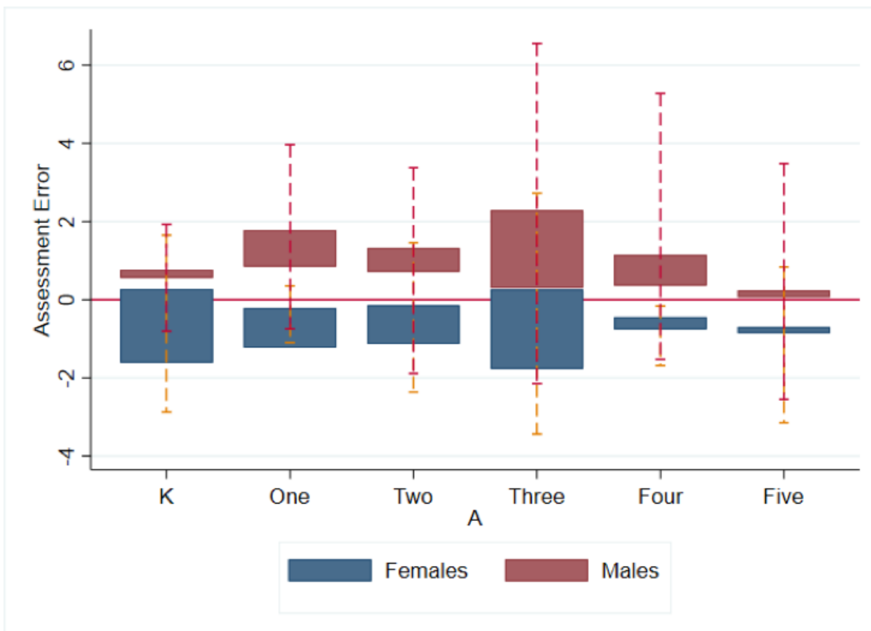
Note: Each point is an average assessment error binned by $days_i$ before or after an eligibility cut-off date that a child's birthday falls, on a range of -180 to 180. Females are plotted in the bottom quadrants in blue and males in the top quadrants in green. Plots use an 1 month bins and a quadratic polynomial fit. Error bars represent 95% confidence intervals, with standard errors clustered on $days_i$.

Figure A4: Special Education Training, by Student Gender



Note: School starting age effects error level of special education training of teacher. Male errors were always above zero and females below zero. Each point shows mean assessment errors binned by $Days_i$ before or after an eligibility cut-off date that a child's birthday falls, on a range of -180 to 180. Plots use an IMSE evenly-spaced optimal bin selection method to calculate bin size and flexible global polynomial fits. Error bars represent 95% confidence intervals, with standard errors clustered on $Days_i$.

Figure A5: Average Treatment Effects Direction by Grade



Note: Each block represents the estimated treatment effect by grade for males (red) and females (blue) from teacher assessed ADHD matched to the estimated intercept point for oldest and youngest males and females in a grade. Each block is estimated using a local linear equation and a two-sided MSE-optimal bandwidth selector with standard errors clustered on $Days_i$. Error bars for 95% confidence intervals on individual intercept points. Results for Grade 6 are omitted due to wide error bars limiting the ability to interpret school starting age effects in earlier grades.