The Persistent Effect of Being Old for Grade on Social-Emotional Skills*

Estelle (Hyewon) Shin†

Click here for the most recent version.

November 5, 2023

Abstract

I examine the persistence of old-for-grade effects in secondary education on social-emotional skills, measured by self-esteem, friendships, learning approaches, and goal-setting mindsets. Students who are older at school entry, thereby older than their grade peers throughout school life may have a potential advantage in their social-emotional skill development. Utilizing a fuzzy regression discontinuity design and leveraging data from the Seoul Education Longitudinal Study 2010 (SELS) in South Korea, I uncover different patterns for girls and boys. I find that old-for-grade girls consistently exhibit higher self-esteem compared to young-for-grade girls during both middle and high school. I also find suggestive evidence that old-for-grade girls maintain more effective learning approaches and stronger friendships. In contrast, I do not find significant differences between old-for-grade and young-for-grade boys. These findings suggest the persistence of old-for-grade effects and gender disparities in social-emotional skill development, potentially carrying implications for future labor market outcomes.

Keywords: School Starting Age, Social-Emotional Skills, Gender Differences, Seoul Education Longitudinal Study 2010 (SELS 2010)

---

*I am very grateful to my advisor, Scott Carrell, for his invaluable advice and continuous support. I also deeply appreciate Marianne Bitler and Jenna Stearns for their guidance and feedback. In addition, I am thankful for the participants at the UC Davis EqualiTEA and UC Davis Applied Micro Student Talks.

†Department of Economics, University of California, Davis. E-mail: heshin@ucdavis.edu

1 Introduction

Social-emotional skills are a crucial facet of human capital that affects individual prospects in the labor market. Researchers have demonstrated that cognitive skills alone cannot fully explain disparities in labor market outcomes (Heckman et al., 2006; Lindqvist and Vestman, 2011). Social-emotional skills predict labor market outcomes including employment, wages, and occupational choices.[1] While a growing body of literature in Economics delves into the impact of schools (Jackson et al., 2020), teachers (Gong et al., 2018; Kraft, 2019), and peers (Gong et al., 2021) on social-emotional skills, there remains a considerable gap in our understanding of the factors that shape these skills.

In this study, I examine whether the effects of being old for grade on social-emotional skills persist throughout middle school and high school. Being old for grade refers to students who are older when they start school, thereby remaining older than their grade peers throughout their school life. As students develop social-emotional skills in school, these developments are influenced by their ages within the same grade. Potential disparities are less problematic if the gap in social-emotional skills between old-for-grade and young-for-grade students naturally fades out over time. However, if these effects persist over time, it becomes necessary to take into account the disadvantages faced by young-for-grade students when designing educational programs aimed at enhancing social-emotional skills.

Students who start school at an older age often possess advantages in cognitive, physical, social, and emotional abilities.[2] However, as students progress into adolescence, it remains

[1] Examples include noncognitive measures combining locus of control and self-esteem (Heckman et al., 2006), social skills (Deming, 2017; Weinberger, 2014), interpersonal skills (Borghans et al., 2008).

[2] Students who start at an older age are more ready for kindergarten (Dhuey et al., 2019), achieve better test scores (Datar, 2006; Elder and Lubotsky, 2009; Cascio and Schanzenbach, 2016), and exhibit fewer problem behaviors and pro-social skills (Datar and Gottfried, 2015; Lubotsky and Kaestner, 2016) during

uncertain how these age-related effects change. On one hand, the observed differences in skill levels between old-for-grade and young-for-grade students fade out if the differences are mainly attributable to a one-year age gap at the same grade. On the other hand, the initial advantages may persistently influence skill development throughout students' education journeys. Old-for-grade students may engage with educational environments that foster social-emotional skill development more effectively (dynamic complementarity).

Previous studies find that the positive effects on approaches to learning and interpersonal skills tend to diminish by the eighth grade in the United States (Datar and Gottfried, 2015; Lubotsky and Kaestner, 2016). Conversely, Crawford et al. (2014) finds no significant effect on self-esteem at age 8, while Datar and Gottfried (2015) finds a higher self-concept among older entrants in the eighth grade, suggesting that not all effects necessarily fade out. By focusing on middle school and high school years, this study provides a deeper understanding of whether the effects of being old for grade on social-emotional skills, measured by self-esteem, friendship, learning approaches, and goal-setting mindsets, persist or fade out.

To investigate this question, I use data from the Seoul Education Longitudinal Study 2010 (SELS), leveraging a unique educational context in Seoul, South Korea. This setting offers several advantages: random middle school assignments, minimal grade repetition, and compulsory middle school. First, the random school assignments, which are uncommon in many other settings, circumvent potential biases in the estimates from sorting into better schools. If old-for-grade outperform their younger peers and are sorted into higher-quality schools, this could change the development of social-emotional skills.[3] Second, the rare

kindergarten.

[3]High school assignment is based on preference in Seoul. Nevertheless, the consistency of the estimates across both middle school and high school indicates that this preference-based assignment does not significantly alter the effects of being old-for-grade in high school. The estimates are also robust to the inclusion

repetition of grades prevents the potential biases that might arise if young-for-grade students are more likely to have lower test scores, resulting in grade repetition and its differential effect on social-emotional skills. Third, compulsory middle school education, with a substantial rate of advancement to high school, ensures that school starting age does not confound the lengths of schooling. If compulsory education is based on the minimum school leaving age, schooling duration can be affected by school starting age, leading old-for-grade students to receive less education.

I employ a fuzzy regression discontinuity design, instrumenting being old for grade with being born after the school entry age cutoff, following the approach used in previous studies.[4] This is necessary because delayed entry is common in my study period, so that decision of school starting age is endogenous. This is less likely to violate the monotonicity assumption, which must hold to interpret the estimates as local average treatment effects and essentially means that students born around the cutoff should be affected in the same direction – becoming older at school entry. Thus, I can interpret my estimates as local average treatment effects.

I find that old-for-grade students exhibit higher levels of social-emotional skills compared to young-for-grade students during middle school and high school. These effects are concentrated among girls. Girls who are old for grade have higher self-esteem than their younger counterparts during both middle and high school. Additionally, there is also suggestive evidence for friendships and learning approaches. Old-for-grade girls exhibit better friendships, indicating stronger social and interpersonal abilities in their relationships with peers, and

of school fixed effects.

[4] Elder (2010); Evans et al. (2010); Landersø et al. (2017, 2020); Johansen (2021)

better learning approaches, indicating increased motivation, engagement, and effective study skills. Notably, the magnitude of these effects during high school is similar to during middle school, suggesting the persistence of the old-for-grade effects. In contrast, there are no significant differences in the social-emotional skills of boys who are old for grade compared to their younger counterparts, except for suggestive evidence of better friendships during middle school. Importantly, these gender-related effects cannot be attributed to differences in contemporary test scores or physical development.

Next, I explore one potential channel for the fadeout of effects. Young-for-grade students may naturally exhibit lower levels of social-emotional skills simply because they are one year younger at the time of the survey. To investigate this possibility, I reconstruct the outcome variables to be measured at the same age. This way, young-for-grade students are at the same age as old-for-grade counterparts and have an additional year of education. If the age at the survey is the primary driver of the effect, or if the extra year of schooling compensates for the advantages of being one year older at the time of the survey, then the effect should be close to zero. I find that the effects are slightly smaller but remain positive and statistically significant when using outcomes measured at the same age. This suggests that the one-year age difference at the time of the survey and one additional year of schooling only partially mitigates the effect of being old for grade on social-emotional skills.

In terms of magnitudes, the old-for-grade effects on social-emotional skills appear to be larger than the effects of high school or peer gender on similar outcomes. For instance, the effects on girls' friendships, ranging from 0.683 to 0.964 standard deviations, align with the effect of increasing the proportion of female peers in the classroom from zero to one (Gong et al., 2021). Additionally, the effects on girls' learning approaches, ranging from 0.536 to 0.916

standard deviations, closely mirror the effect of school starting age in kindergarten, which is 0.701 standard deviation according to Lubotsky and Kaestner (2016). These substantial effect sizes may, in part, be attributed to the local nature of both regression discontinuity and instrumental variable designs.

My findings hold important implications for post-secondary education and the labor market prospects, particularly for old-for-grade girls. While I cannot directly observe students after their high school graduation, previous studies indicate that enhanced social-emotional skills play a vital role in labor market outcomes. Back-of-the-envelope calculations suggest that a one standard deviation increase in self-esteem is associated with approximately a 25.06 percent increase in wages for females (de Araujo and Lagos, 2013), and a one standard deviation increase in social skills is associated with a 3.0 percentage points increase in full-time employment across genders (Deming, 2017). Moreover, higher levels of learning approaches have the potential to enhance post-secondary educational attainment. In summary, the improved social-emotional skills among old-for-grade girls provide a promising path towards enhanced labor market outcomes. Given the importance of social-emotional skills, it is crucial to consider the potential disadvantages faced by young-for-grade students when developing programs aimed at fostering social-emotional development.

There are several contributions of my paper. First, by focusing on middle school and high school, I demonstrate the persistence of being old-for-grade effects in secondary education. The inclusion of high school in the analysis extends beyond previous research, allowing for the observation of these effects closer to labor market entry. Second, I uncover distinctive patterns in the development of social-emotional skills among girls and boys, providing evidence that old-for-grade girls persistently experience positive effects on their social-emotional skills during adolescence, whereas such trends are not observed among boys. Third, I analyze whether one-year age differences can account for the skill gap measured at the same grade. My findings suggest that these differences do not entirely explain this gap. Lastly, my study provides robust causal evidence regarding the effects of being old for grade, leveraging the random assignment of middle schools, rare grade repetition, and fixed years of compulsory education, and employing an instrument that satisfies necessary identifying assumptions.

The remainder of this paper is organized as follows. Section 2 describes the educational system in South Korea and the data I use. Section 3 explains the identification strategy, and Section 4 presents the results and interpretation. Section 5 presents robustness checks. Section 5 concludes.

2 Educational System and Data

Educational System in South Korea

Education in South Korea consists of six years of primary school, three years of middle school, and three years of high school. The academic year begins in March, and the school starting age cutoff date was March 1st until 2008.[5] Children become eligible to enter primary school in March following their 6th birthday. Consequently, if children enter school on time, March-born children turn 7 years old, while February-born children are 6 years old at school entry. Although early school entry or delayed entry is possible, early entry is relatively rare,

[5]The school starting age cutoff moved from March 1st to January 1st in 2008, and the January 1st cutoff was enforced starting in 2009. Children who turn six years old by January 1st are eligible to enter primary school in March of that year, effectively starting school in the calendar year of their 7th birthday. The sample utilized in this paper entered school when March 1st was the school starting age cutoff.

with only about 0.3 - 0.6 percent of children in Seoul entering school early each year.[6] On the other hand, delayed school entry is more common, with approximately 5.4 percent in 2004 and about 8.7 percent in 2007 delaying their school entry in Seoul.[7] [8]

Before entering primary school, children in South Korea typically attend either kindergarten or daycare centers. Kindergarten is not entirely free when the children in my sample are aged 3 to 5, and kindergarten and daycare centers did not have national curriculum before 2012. The share of children attending either kindergarten or daycare centers at age 5 was 76.3 percent in 2004 and 87.9 percent in 2007 in Korea. Most children are enrolled in one of these programs, which indicates that many old-for-grade children have one more year of experience in pre-primary education before school entry. This also indicates that old-for-grade and young-for-grade children experience different education cycles at the same age.[9]

Grade repetition is infrequent in South Korea. In 2010, only a small number of students, 12 in primary school and 45 in middle school, repeated grades.[10] Grade repetition may have differential impacts on the development of social-emotional skills compared to students advancing to higher grades. Additionally, young-for-grade students are more likely to have

[6] In population-level data in Seoul, 779 out of 124,209 children entered school early in 2004, and 377 out of 111,496 students entered early in 2007. The share of early entrants remained relatively stable until the school entry cutoff moved to January 1st. Source: Statistical yearbook of Seoul education 2005, 2008

[7] To delay school entry, parents need to file an application for reason such as child development, and the primary school headmaster decide whether to accept the application following the deliberation of the compulsory education management committee.

[8] Nationwide delayed entry rates in the United States are similar to those in Korea. In the United States, delayed entry rates range from 5 percent to 7 percent in recent years (Datar, 2006; Bassok and Reardon, 2013; Snyder and Dillow, 2013). Delayed entry in the United States was more prevalent in the 1970s and 1980s, with an average of 12 percent, but has declined over time (Huang, 2015).

[9] Kindergarten provides care for children from age 3 until they transition to primary school, while daycare centers provide care for children aged 0-6. These two types of early childhood education and care facilities operate under different laws and government supervision. Among children aged 5 in South Korea who attend kindergarten or daycare centers, approximately 60 percent attended kindergarten, while the remaining attended daycare centers.

[10] Source: Statistical yearbook of education

lower test scores, leading to a higher likelihood of repeating a grade (Dhuey et al., 2019). This suggests that grade repetition may result in differential social-emotional development trajectories for young-for-grade students. In my setting, the infrequency of grade repetition ensures unbiased estimates.

South Korea mandates six years of primary school and three years of middle school. 99.7 percent of middle school graduates advance to high school with an annual high school dropout rate of only one percent.[10] While the minimum school leaving age can potentially influence the length of schooling based on when students start school (Barua and Lang, 2016), the combination of nine years of compulsory education and a substantial advancement rate to high school reduces concerns related to dropout and its impact on social-emotional development.

In Seoul, middle school assignments are random within 11 school districts. High school assignments are based on student preferences. Most students choose between academic and vocational high schools. Within academic high schools, students submit preferences. Random school assignments in middle schools prevent my estimates from being confounded by school quality effects as old-for-grade students are more likely to outperform young-for-grade students in primary school.

Data

I use the Seoul Educational Longitudinal Survey 2010 (henceforth SELS 2010) for this study. SELS 2010 samples fourth, seventh, and tenth grade students in 2010 and conduct annual surveys until their high school graduation.[11] I utilize data from the fourth grade and seventh grade cohorts to analyze the effects in both their high school years (grades 10 to 12) and middle school years (grades 7 to 9). After dropping students with missing dates of birth, there are 3,834 students in the fourth grade cohort and 4,515 students in the seventh grade cohort.[12] To address potential concerns regarding non-random attrition in the panel survey, I investigate whether attrition is correlated to being born after March 1st or being old-for-grade. However, the correlation between attrition and being born after the cutoff or being old for grade is close to zero, suggesting that attrition is unlikely to significantly bias the results.

One limitation of this data set is that it is not possible to directly observe the school starting age. Instead, I use the observed age in the first survey wave as a proxy for the school starting age. If students have repeated a grade or experienced grade retention, the observed age may not perfectly align with the actual school starting age. However, this is less problematic due to the infrequency of repetition or retention in South Korea. Retention typically occurs when a student's attendance rate falls below two-thirds of the total school days.In 2007, only 9 primary school and 23 middle school students repeated grades in Seoul. Furthermore, students can postpone school enrollment, but this is only allowed for reasons such as illness or other extenuating circumstances.[13] n this sample, the majority of students

[11]SELS 2010 utilizes a stratified two-stage cluster sampling approach. In the first stage, schools are randomly selected based on the number of schools within each school district. In the second stage, two classes are randomly chosen within each school.

[12]Some students are added in survey waves 6 and 7. 164 students in the fourth grade cohort and 33 students in the seventh grade cohort are added for a new policy evaluation in the sixth survey wave. 75 students are additionally added for the fourth grade cohort in the seventh survey wave.

[13]Elementary and Secondary Education Act Article 28 Paragraph (4): The head of an elementary school or a middle school shall grant the exemption or postponement of the obligation of school enrollment pursuant to paragraph (2) or (3) only where the application thereof is filed because of illness or any extenuating circumstances determined by the superintendent of education.

in the seventh grade cohort were born in 1997 and 1998, while the majority in the fourth grade cohort were born in 2000 and 2001. Only a small number of students are older or younger than their assigned school age, which reassures that the observed age in the fourth or seventh grade likely reflects the actual school starting age.

The main outcomes of interest are self-esteem, friendship, learning approaches, and goal-setting mindsets. These social-emotional skills have been demonstrated to have an effect on academic attainment and labor market outcomes in previous studies. For instance, studies have shown that self-esteem affects educational attainment, wage, employment, and occupational choices (Heckman et al., 2006; de Araujo and Lagos, 2013). The friendship index measures a student's social and interpersonal skills, which predict later employment and occupational choices (Deming, 2017; Lindqvist and Vestman, 2011; Borghans et al., 2008). The learning approaches index measures students' learning strategies, effort, and motivations, all of which can positively affect their educational attainment. Finally, the goal-setting mindsets index measures whether students have clear goals, know how to achieve them, and work hard to attain them, indicating a connection to motivation.

The self-esteem index includes five questions from the Rosenberg Self-Esteem scale, which is widely used in Economics research. These questions are: *I feel that I have a number of good qualities. I am able to do things as well as most other people. I feel that I'm a person of worth. I take a positive attitude toward myself. On the whole, I am satisfied with myself.*

The friendship index includes four questions of the following: *I have a friend whom I can trust and talk to. Rather than being alone during breaks or lunchtime, I spend time with my friends. Even if I have arguments with my friends, we make up quickly. I help my friends* *who need assistance.*[14]

The self-directed learning approach index includes nine questions based on the motivated strategies for learning questionnaire (MSLQ) (Pintrich et al., 1993), which include: *I think about how to connect newly learned information with what I already know. I study important materials by summarizing key points or organizing them in tables or mind maps. I check myself to see if I understand the class materials well. I try my best to fully understand the school materials. I try to stick to the study schedule I planned as much as possible. If there is anything I don't understand while studying or doing assignments, I look it up in books or online. I believe I will eventually understand any difficult material. I find studying enjoyable. I can handle my own work well without anyone telling me what to do.*[15]

The goal-setting mindset index includes five questions: *I have a clear goal that I want to achieve. I know what I need to do to achieve my goals. I am working hard to achieve my goals. The study I am doing now will help me achieve my future goals. If I achieve my future goals, I believe I can also contribute to society.*

Students respond to each question using a 5-point Likert scale, ranging from 1 being 'strongly disagree' to 5 being 'strongly agree.'[16] A higher score means better social-emotional skills. I present summary statistics for outcome indices by gender and cutoff among students born between January and April in Table 1. Across outcomes and grades, girls born in March or April, and thus more likely to be older, have higher average indices than those born in January or February. On the other hand, for boys, being born after March 1st does

[14]Some of these questions are similar to the interpersonal skill survey questions in Jackson et al. (2020) and Datar and Gottfried (2015); Lubotsky and Kaestner (2016).

[15]Questions used to measure the learning approach index are similar to academic effort (work hard index) in Jackson et al. (2020).

[16]Cronbach's α is reported for the reliability of these measures in the fourth survey wave. Most measures score above 0.7.See Appendix Table A18 for detail.

not consistently result in higher average indices. Between genders, girls have slightly lower self-esteem than boys, but there is no consistent pattern of girls scoring higher or lower than boys in other social-emotional skills.

To better interpret estimates, I standardize each survey response to mean zero and unit variance within each survey wave and sum relevant questions. I average each outcome over middle school years (grades 7 to 9), high school years (grades 10 to 12), across all grades (grades 7 to 12), and across all ages (ages 13 to 17) to reduce noise due to a small sample size.[17] All outcomes are then standardized again. I use ages 13 to 17 when averaging across ages because data is unavailable for young-for-grade students at age 18 and old-for-grade students at age 12.[18] Young-for-grade students are in grade 8 at age 13 and grade 12 at age 17, while old-for-grade students are in grade 7 at age 13 and in grade 11 at age 17.

3 Identification Strategy

I use a fuzzy regression discontinuity design to investigate the old-for-grade effects on social-emotional skills. Despite the exogenous nature of dates of birth around the school entry cutoff, the timing of a child's school entry remains endogenous. Parents may choose to delay their child's school entry if they believe the child is not sufficiently ready for school. Therefore, I adopt being born after the school entry cutoff as an instrumental variable for the binary variable old-for-grade, similar to previous studies such as Evans et al. (2010) and Johansen (2021).

[17] I also estimate the effects on outcomes at a specific grade or a specific age, but these estimates are consistent with those using pooled averages.

[18] The survey was conducted in July every year, so the age is calculated based on the survey time.

Old-for-grade is a binary variable, equal to one if the school starting age is 7 and zero if the school starting age is 6. Figure 1 illustrates the share of old-for-grade children by gender.[19] Children born in January and February are more inclined to delay their school entry as their dates of birth approach the cutoff. There is a jump in the share of old-for-grade children on March 1st, with the majority of children born in March and April being old for grade. Consistent with prior research (Fredriksson and Öckert, 2014; Landersø et al., 2017; Cook and Kang, 2020), boys are more likely to delay school entry compared to girls.

I estimate the following equations in the reduced form and in the first stage:

$$Y_{ic} = \beta_1 + \beta_2 cutoff_i + \beta_3 d_i + \beta_4 d_i * cutoff_i + \lambda_c + X_i'\gamma_1 + \epsilon_i$$

$$oldforgrade_{ic} = \alpha_1 + \alpha_2 cutoff_i + \alpha_3 d_i + \alpha_4 d_i * cutoff_i + \delta_c + X_i'\gamma_2 + e_i$$

where Y_i is the outcome of interest, which includes self-esteem, friendship, learning approaches, and goal-setting mindsets in middle school and high school. d_i is the date of birth relative to the March 1st cutoff. It is equal to 0 when a child is born on March 1st, 1 if born on March 2nd, and -1 if born on February 28th. $cutoff_i$ is a binary variable equal to one if children are born after March 1st. I also include an interaction between d_i and $cutoff_i$. λ_c and δ_c are the survey cohort fixed effects in the reduced form and the first stage respectively. X_i is predetermined individual characteristics, including gender, firstborn, parents' educational attainment (categorized as less than high school, high school graduates, and more than high school), and indicators for missing covariates. I do not include individual controls in my main specification because the estimates remain robust to the inclusion of controls.

[19] The share of old-for-grade students can be interpreted as old-for-grade within their respective grade cohorts in Seoul.

The estimate of interest is hence $\hat{\beta}_2/\hat{\alpha}_2$.

I use a triangular kernel, which is optimal for boundary estimation, using local linear regression (Fan and Gijbels, 1996). In the main analysis, I choose 60-day bandwidths to ensure comparison of the same sample across outcomes. Moreover, I conduct additional analyses using bias-corrected estimates and robust confidence intervals, utilizing bandwidths that minimize the mean squared errors (MSE) following Calonico et al. (2014). The results remain robust. Results are shown in Appendix Table A13. To further validate my approach, I investigate the optimal local polynomial order that minimizes asymptotic mean squared errors, following Pei et al. (2022). I find that local linear estimators consistently have lower asymptotic mean squared errors than local quadratic or cubic estimators. Results are shown in Appendix Tables A2 and A3.

I cluster the standard errors at the date of birth level. Additionally, I consider alternative clustering at the school level, given that SELS 2010 is sampled at the school level, and students within the same school might be exposed to common shock. However, the alternative clustering does not yield significantly different results compared to my main specification. Results are shown in Appendix Table 6. Notably, standard errors are larger when clustering at the date of birth level. This suggests that clustering at the date of birth level provides more conservative results. Therefore, in my main results, I present standard errors clustered at the date of birth level for robust and conservative estimates.

The validity of the regression discontinuity design relies on the assumption of continuous running variables and the smoothness of the regression functions for potential treatment and potential outcomes near the cutoff. One possible threat to identification is that parents may manipulate the date of birth around the school entry cutoff (Shigeoka, 2015). To address this, I first check whether the density of the running variable around the cutoff changes smoothly (McCrary, 2008). Figure 2 presents the density of births within a 60-day bandwidth around the cutoff with two-day bins. The number of births is not significantly different before and after the cutoff. Second, I formally test the manipulation using the method proposed by Cattaneo et al. (2020). Results shown in Figure 3 do not reject that the density around the cutoff is smooth. I also do the same analysis separately for girls and boys in the Appendix Figures A1 and A2.[20] Lastly, I re-estimate the effects with the donut regression discontinuity design in the robustness check. I exclude one, four, and seven days around the cutoff. Results are shown in Appendix Figures A9 and A10. While the magnitudes remain largely consistent, but they become somewhat noisy.

Another potential source of bias in the density of birth is the possibility of unique birth patterns in two specific school cohorts, since I use only a sample of students from these cohorts in my data. To mitigate this concern, I analyze whether residuals, obtained by regressing running variables on day-of-week and holiday fixed effects, are smooth around the cutoff. The estimates, shown in Appendix Figure A3, are robust to the inclusion of day-of-week fixed effects. Furthermore, Kim (2021) finds that adjustments in birth timing between February and March from 1997 to 2007 are small and not statistically significant, based on administrative birth certificates in Korea.[21] This evidence suggests that the birth patterns

[20] One possible concern related to my data is that the distribution of births around the cutoff date might differ from that at the population level. If, for instance, time-specific events impact the two school cohorts in my data, my estimates might be biased. While I do not have access to the daily number of births at the population level in Seoul, I find that the share of births by month in my data is similar to that at the population level in Seoul. This evidence assures that the cohorts in my sample are not drastically different from the overall population in terms of birth patterns.

[21] One explanation provided by Kim (2021) is that school starting age enforcement is flexible, so parents can adjust the timing of their child's school entry when the child is eligible to enter school. This gives less incentive for parents to adjust birth timing.

in my sample are not atypical compared to other cohorts and the population at large.

Another way to assess the validity of regression discontinuity design is to check whether individual predetermined characteristics are balanced around the cutoff. Figure 4 presents the distribution of females, firstborns, and parents' educational attainment for students born between January and April. The bin size is set to two days. Overall, the predetermined characteristics of children appear to be similar before and after the March 1st cutoff, suggesting no significant imbalance in these characteristics. One exception is that there is a slightly greater proportion of girls among January and February-born children compared to those born in March and April. I present the results separately for girls and boys, in addition to the entire sample.

Since I am using the instrument, standard instrumental variable assumptions, including the relevance of the instrumental variable and the exclusion restriction, must also be satisfied. Additionally, the monotonicity assumption is necessary to evaluate the local average treatment effect (LATE). Monotonicity implies that students born around the cutoff, if they are affected, should be affected in the same direction—specifically, they should become old-for-grade at school entry if affected in my setting. Instrument variables used in previous studies such as legal school entry age may violate the monotonicity assumption (Aliprantis, 2012; Barua and Lang, 2016; Fiorini and Stevens, 2021). Because I instrument a binary variable old-for-grade with being born after the cutoff, 'defiers' refer to parents and children who delay school entry when children are born before the cutoff and enter early when children are born after the cutoff. In other words, children will be age 7 when they are born before the cutoff and age 6 when they are born after the cutoff. This seems implausible. For example, if children are not ready for school at age 6, they may choose to be old-for-grade regardless of whether they are born before or after the cutoff. On the other hand, if children are already ready for school at age 6, they will enter school on time if they are born before the cutoff. It is possible that they will enter early or on time if they are born after the cutoff. Hence, it is not likely that the monotonicity assumption is violated.

4 Results

First Stage

I present the first-stage regression results in Table 2. All regressions include cohort fixed effects. Columns 1 and 2 show estimates for the entire sample. Columns 3 and 4 show estimates for girls, and Columns 5 and 6 show estimates for boys. The results show that children born after the cutoff are 32.3 percentage points more likely to be old-for-grade in their grade cohort. The effects are larger for girls with 37.6 percentage points compared to boys with 27.8 percentage points. Given that 44.6 percent of girls and 55.3 percent of boys born in January and February delay their entry, boys are more likely to be old for grade when they are born before cutoff. The estimates are significant in all regressions and remain consistent with and without individual controls. In the following results, I only present regression results without controls.

Main Results

Table 3 presents the main estimation results for social-emotional skills during middle school and high school. Odd columns present the effects during middle school (grades 7 to 9), while even columns present the effects during high school (grades 10 to 12). Panel A illustrates the results for girls, Panel B for boys, and Panel C for the entire sample. Columns 1 and 2 show the effects of being old for grade on self-esteem. Girls who are old for grade are more likely to exhibit higher self-esteem compared to young-for-grade girls in both middle school and high school. Magnitudes are similar in both periods as well, suggesting persistence of the effects. On the other hand, the effect on self-esteem for old-for-grade boys is relatively small and statistically insignificant. Panel C suggests that students who are old for grade generally exhibit higher self-esteem. However, Panel A and B indicate these effects are primarily driven by girls.

The effects on friendships, in comparison to those on self-esteem, are generally less pronounced. On the whole, students who are old for grade tend to establish more positive relationships with their peers when compared to their young-for-grade counterparts within the entire sample. Old-for-grade females, again, persistently enjoy more robust friendships during both middle and high school. In the case of males, the estimate on friendships is not statistically significant, despite the effect size resembling that observed in females during middle school. Intriguingly, as students progress to high school, the distinction between old-for-grade and young-for-grade males becomes notably diminished.

Columns 3 and 4 presents the effects on friendships. In comparison to those on self-esteem, the effects are generally less pronounced. Overall, old-for-grade students tend to have better social and interpersonal skills in the relationships with their peers when compared to their young-for-grade counterparts within the entire sample. Old-for-grade females persistently enjoy better friendships during both middle and high school. For boys, however, the estimate on friendships is not statistically significant, despite the similar effect size to girls during middle school. Interestingly, as students progress to high school, the difference between old-for-grade and young-for-grade boys fades out. Note that the evidence for the positive effect on friendships should be taken cautiously, given the sensitivity of the estimates to various specifications and bandwidth choices. Nonetheless, the findings imply that old-for-grade girls experience persistently better friendships, while the positive effects on old-for-grade boys tend to diminish.

Columns 5 and 6 present the effects of being old for grade on learning approaches. Girls who are old for grade tend to exhibit more effective study habits, greater engagement in learning, and heightened dedication when compared to their younger counterparts. Again, these effects persist throughout middle and high school. The estimate of 0.916 standard deviations in high school slightly bigger than the estimate in middle school. One plausible explanation could be the higher perceived stakes associated with high school, as test scores in this period directly influence college admissions. Consequently, old-for-grade girls might exhibit enhanced self-control and study behaviors.

Panel A of columns 7 and 8 suggests that old-for-grade girls tend to demonstrate a better ability to set goals. This implies that they are more likely to have clear goals and actively strive to achieve them. However, the estimates are somewhat smaller and statistically significant only at the 10 percent level during middle school, eventually losing their statistical significance in high school. In contrast, the effects of learning approaches and goal-setting mindsets for males are relatively negligible during middle school, with a shift towards a negative direction in high school, albeit without statistical significance. Taken together, old-for-grade boys are not significantly more likely to have better learning approaches or goal-setting mindsets compared to their younger counterparts.

In summary, the old-for-grade effects are more pronounced for girls. Old-for-grade girls exhibit higher self-esteem, stronger relationships with friends, and more effective learning approaches. Importantly, these findings suggest that the positive effects for old-for-grade girls persist over time rather than fade out as the students grow older. In contrast, old-for-grade boys do not exhibit noticeable improvements in their social-emotional skill development.

To provide a broader context in terms of magnitudes, I first compare my estimates to other papers that study the effect of school starting age on similar outcomes. The effect size for self-esteem, ranging from 0.590 to 0.762 standard deviations for the entire sample, surpasses the 0.15 standard deviation identified by Datar and Gottfried (2015) for eighth graders. Notably, the impact on learning approaches for old-for-grade girls does not diminish over time, in contrast to the findings of Datar and Gottfried (2015) and Lubotsky and Kaestner (2016). The magnitude of the estimate in high school is comparable to the 0.63 standard deviation observed in Lubotsky and Kaestner (2016) or the 0.701 standard deviation identified in Datar and Gottfried (2015) during the Fall of kindergarten.

Then I compare my estimates with other interventions examining similar outcomes. This difference can potentially be attributed to variations in estimation methodologies and the local nature of both regression discontinuity and instrumental variable approaches. Despite potential limitations in terms of external validity beyond the specified bandwidths, reduced-form estimates could be employed for comparison with other interventions. Estimation results from reduced-form regressions are presented in Appendix Table A7. For instance, the effect on girls' friendships, ranging from 0.257 to 0.372 standard deviations, is comparable in magnitude to the effect of increasing the proportion of female peers in the classroom by 33.5 to 48.5 percentage points Gong et al. (2021). This increases students' social acclimation and general satisfaction index. Similarly, the effect on the learning approaches index of 0.353 standard deviation in high school aligns with the magnitude associated with an increase in teacher quality by 2 to 3 standard deviations on students' effort in class (Kraft, 2019).

Since there are multiple outcomes across middle school and high school, I also investigate the possibility of an increased number of significant results due to multiple hypothesis testing. I employ Bonferroni and Sidak corrections to adjust the p-values. I first adjust p-values across outcomes for girls. While the impact on friendship and learning approaches is not robust, the difference in self-esteem between old-for-grade girls and young-for-grade girls remains statistically significant. Furthermore, girls' self-esteem remain significant even after implementing corrections across both gender and outcomes. These findings further support that old-for-grade girls are more likely to exhibit higher self-esteem than young-for-grade girls, although the effects on other social-emotional skills might not be as robust.

Mechanisms and Interpretation

An important question arises: does this effect fade out over time due to absolute age differences at the time of the survey? It is plausible that young-for-grade students exhibit lower levels of social-emotional skills because they are one year younger at the same grade. Then, young-for-grade students might catch up when they reach the same age as old-for-grade students. To investigate this possibility, I compare estimates using outcome measures at the same grade with estimates employing outcomes measured at the same age. The former include the combined effect of school starting age, relative age compared to peers, and age at survey. The latter control for the age at the survey but may include a potentially negative effect related to having one less year of schooling for old-for-grade students. For instance, at the age of 17, old-for-grade students are in grade 11, while young-for-grade students are in grade 12. Hence, old-for-grade students are at the same age, but also have one less year of schooling compared to young-for-grade students. If age at survey drives the entire effect, the estimates using the outcomes measured at the same age should be smaller and not statistically significant.

Ideally, it would be best to use social-emotional skill measurements in the later stages of life to fully control for years of schooling and differences in educational cycles. However, the data I use do not include information beyond high school. Nonetheless, I can compare the estimates at all grades (grades 7 to 12) and those at all ages (ages 13 to 17) to partially examine the potential influence of years of schooling.

Table 4 presents the results. For girls, I find that the estimates at all ages are slightly attenuated compared to those at all grades. This suggests that having one more year of schooling might partially mitigate the effect of starting school at an older age or that the age at the survey contributes to some of the estimated effects in middle and high school years. The effects for boys are generally consistent across all grades and all ages as well.[22] This indicates that the old-for-grade effects do not completely fade out when young-for-grade students are at the same age as old-for-grade students, and one more year of schooling does not completely compensate for the starting school at an older age. Age at school entry and the relative age compared to peers play a role in the social-emotional skill development.

Next, I investigate the potential mechanisms that might explain these effects. First, I as-

[22] These findings remain consistent when I compare the estimates of skills measured at a specific grade and those measured at a specific age. Results are shown in Appendix Tables A8 to A11.

sess whether old-for-grade students have higher test scores.[23] Though imprecisely estimated, old-for-grade girls experience positive effects on test scores in middle school. However, the estimates tend to converge toward zero in high school. This suggests that the disparities in girls' self-esteem cannot be solely attributed to differences in contemporaneous test performance.

The lack of significant effect on test scores, despite the positive impact on learning approaches, may seem counterintuitive. One plausible explanation is that the exams in the survey may not carry as high stakes. In fact, Kim (2011) finds that a one month increase in school starting age increases the likelihood of entering a 4-year college by 0.42 percentage points within two years after high school graduation in Korea. This suggests that, although the direct influence on test scores may not be immediately evident, the effects of being old for grade on learning approaches could still play a role in later life.

Second, I investigate whether old-for-grade students possess superior physical abilities. Due to data constraints, I analyze the height and weight of each cohort in the first survey wave. The estimates for girls in the fourth-grade cohort are positive, albeit not statistically significant, while the estimates for boys are small and lack statistical significance as well. The estimates within the seventh-grade cohort are smaller than those in the fourth-grade cohort. It appears that old-for-grade girls might be taller than their younger peers in the fourth grade, but this difference diminishes by seventh grade. It is plausible that their early physical advantages influence self-esteem before entering middle school and persist throughout middle and high school.

[23]Test scores are derived from Korean, math, and English tests developed by SELS 2010. Questions are based on the first-semester curriculum of each grade in the survey year.

Improved social-emotional skills for old-for-grade girls have crucial implications. Self-esteem has positively correlated with post-secondary education and later labor market outcomes, including employment and earnings (de Araujo and Lagos, 2013). Based on the estimates provided by de Araujo and Lagos (2013), a back-of-the-envelope calculation suggests that a one standard deviation increase in self-esteem could result in an approximately 25.06 percent wage increase for females, through an increase in years of schooling. Friendships, which are combined index for social and interpersonal skills, can similarly impact employment. According to Deming (2017), a one standard deviation improvement in social skills predicts a 3.0 percentage point rise in full-time employment. Given the 0.964 standard deviation increase in friendship for girls, this could potentially translate into a 2.89 percentage point increase in female employment. While interpreting these findings, it is essential to be cautious since most previous studies are based on US data, and some date back to the 1980s and early 2000s, which may differ from the current South Korean context. Nevertheless, it remains plausible that old-for-grade girls with enhanced social-emotional skills might attain more years of education and experience improved labor market outcomes in the future.

5 Robustness Check

In this section, I establish the robustness of the results through a series of tests. First, I replicate the analyses with different bandwidths. Specifically, I employ mean squared error (MSE) optimal bandwidths for individual regressions and the first stage, both computed following Calonico et al. (2014). I compare bias-corrected estimates and robust confidence intervals with the main estimates. Secondly, I explore the sensitivity of the results to different functional forms by estimating the effects using global linear and quadratic regressions, also utilizing various bandwidths (30, 60, and 90 days). Third, I assess whether a donut regression discontinuity approach, which excludes a small number of observations close to the cutoff, yields different results. Fourth, I examine whether including day-of-week and holiday fixed effects in the specification affects the main results.

To evaluate the sensitivity to bandwidth choices, I re-estimate the effects using MSE optimal bandwidths, following Calonico et al. (2014). I employ two approaches for computing MSE optimal bandwidths. The first minimizes the mean squared errors for fuzzy regression discontinuity design, generating different bandwidths for each regression. The MSE optimal bandwidths for fuzzy regression discontinuity range from 42 days to 78 days.The second uses the MSE optimal bandwidths computed in the first stage. Thus, I use the same bandwidths within each gender. The first-stage MSE optimal bandwidths are 38 days for girls and 40 days for boys. Results are presented in the Appendix Figures A5 and A6. The estimates maintain similar magnitudes across the different bandwidth choices. Estimates also remain statistically significant except for girls' friendships in high school and learning approaches in all grades. This indicates that the results reported earlier are robust and not sensitive to bandwidth choices.

Concerns might arise due to dates of birth being discrete variables, raising questions about the applicability of local linear regression discontinuity design. To address this concern, I replicate the main results using a parametric regression discontinuity design and cluster the standard errors at the date of birth level, following the approach proposed by Lee and Card (2008). The results obtained through global linear polynomials are largely consistent with my main specifications. To further check sensitivity to functional form, I compare estimates using linear and quadratic control functions and different bandwidth choices (30 days and 90 days). Results are shown in the Appendix Figures A7 and A8. Some estimates become noisy when using a smaller bandwidth of 30 days. This might be attributed to a smaller sample size due to narrower bandwidths. Although some estimates lose statistical significance, the magnitudes are consistent in general. Importantly, results for girls' self-esteem remain large and significant across different specifications. This reassures that the effects on girls' self-esteem are robust to changes in functional forms.

Next, I redo the same analyses using a donut regression discontinuity approach. Although manipulation testing does not reject the smoothness of the density of births, one might be worried about the randomness of births around the cutoff, especially considering that March 1st is a holiday in Korea. To examine the robustness of the results, I exclude 1 day, 4 days, and 7 days before and after March 1st. Appendix Figures A9 and A10 indicate that the coefficients are robust to the donut approach. While the effects on girls' friendships and goal-setting mindsets become larger and statistically significant when excluding 7 days before/after the cutoff, the magnitudes of other outcomes generally remain similar.

Additionally, I investigate whether the results are sensitive to including day-of-week and holiday fixed effects. Results are shown in Appendix Tables A13 to A16. Columns 1 and 5 show estimates from my main specifications, and Columns 4 and 8 show estimates with day-of-week fixed effects. The estimates remain nearly unchanged across these different specifications. The results from the donut approach and the inclusion of day-of-week fixed effects help address concerns regarding potential birth manipulation around the cutoff.

6 Conclusion

In this study, I investigate whether the effects of being old for grade on students' social-emotional skills fade out. Using a binary treatment of old-for-grade and a binary assignment of being born after the cutoff date, my identification strategy of fuzzy regression discontinuity design reduce concerns of monotonicity violation. Furthermore, the unique educational setting in Seoul, including random middle school assignment, rare grade repetition, and a substantial high school advancement rate, allow me to isolate the effect of being old-for-grade from potential confounding factors related to school sorting, repetition, and education duration.

Extending the study period to high school, closer to labor market entry, I find that old-for-grade girls persistently exhibit higher self-esteem than young-for-grade girls in both middle school and high school. There is also suggestive evidence that they maintain better friendships throughout adolescence and display better learning approaches in high school. In contrast, the effects on boys' social-emotional skills are relatively limited. Old-for-grade boys do not exhibit significantly higher social-emotional skills than their young-for-grade counterparts. They do show a slight improvement in friendship during middle school, but this effect dissipates in high school. These findings highlight distinct patterns of social-emotional skill development for girls and boys. Lastly, I compare the estimates using outcomes measured at the same age to the estimates using outcomes measured at the same grade. When absolute age at the survey is controlled, old-for-grade girls still exhibit higher social-emotional skills even though they have one less year of schooling.

However, there are several limitations to this study. First, the study lacks information on the ages of peers in the classroom or school beyond the first year of the survey. This prevents me from disentangling the absolute age and relative age effects. Consequently, the estimates only capture the combined effect of being older at school start and relatively older than their peers. Second, the data used in this study only covers the period until students graduate from high school. Therefore, it is not possible to analyze whether the positive effects of being old-for-grade translate into long-term labor market outcomes.

Nonetheless, this study provides evidence that the effects of age for grade on social-emotional skills do not fade out over time. Furthermore, the findings underscore the importance of considering gender and age differences in skill development. The presence of defined cutoff dates within the school system results in students being categorized as either old-for-grade or young-for-grade around the cutoff date. Given the persistent differences between these two groups, it is crucial to address the potential disadvantages faced by young-for-grade students. This research has practical implications for educators, highlighting the importance of designing effective programs to support the development of social-emotional skills in children and adolescents, taking into account gender and age differences within the same grade.

References

Aliprantis, Dionissi, "Redshirting, Compulsory Schooling Laws, and Educational Attainment," *Journal of Educational and Behavioral Statistics*, 2012, *37.*

Barua, Rashmi and Kevin Lang, "School entry, educational attainment, and quarter of birth: A cautionary tale of a local average treatment effect," *Journal of Human Capital*, 2016, *10.*

Bassok, Daphna and Sean F. Reardon, ""Academic Redshirting" in Kindergarten: Prevalence, Patterns, and Implications," *Educational Evaluation and Policy Analysis*, 2013, *35.*

Borghans, Lex, Bas ter Weel, and Bruce A. Weinberg, "Interpersonal Styles and Labor Market Outcomes," *Journal of Human Resources*, 2008, *43*, 815–858.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik, "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 11 2014, *82*, 2295–2326.

Cascio, Elizabeth U. and Diane Whitmore Schanzenbach, "First in the class? Age and the education production function," *Education Finance and Policy*, 2016, *11.*

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma, "Simple Local Polynomial Density Estimators," *Journal of the American Statistical Association*, 7 2020, *115*, 1449–1455.

Cook, Philip J. and Songman Kang, "Girls to the front: How redshirting and test-score gaps are affected by a change in the school-entry cut date," *Economics of Education Review*, 6 2020, *76*, 101968.

Crawford, Claire, Lorraine Dearden, and Ellen Greaves, "The drivers of month-of-birth differences in children's cognitive and non-cognitive skills," *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 2014, *177.*

Datar, Ashlesha, "The impact of kindergarten entrance age policies on the childcare needs of families," *Journal of Policy Analysis and Management*, 12 2006, *25*, 129–153.

_ **and Michael A. Gottfried**, "School entry age and children's social-behavioral skills: evidence from a national longitudinal study of U.S. kindergartners," *Educational Evaluation and Policy Analysis*, 2015, *37.*

de Araujo, Pedro and Stephen Lagos, "Self-esteem, education, and wages revisited," *Journal of Economic Psychology*, 2 2013, *34*, 120–132.

Deming, David J., "The growing importance of social skills in the labor market," *Quarterly Journal of Economics*, 11 2017, *132*, 1593–1640.

Dhuey, Elizabeth, David Figlio, Krzysztof Karbownik, and Jeffrey Roth, "School Starting Age and Cognitive Development," *Journal of Policy Analysis and Management*, 2019, *38.*

Elder, Todd E., "The importance of relative standards in ADHD diagnoses: Evidence based on exact birth dates," *Journal of Health Economics*, 2010, *29.*

_ and Darren H. Lubotsky, "Kindergarten Entrance Age and Children's Achievement: Impacts of State Policies, Family Background, and Peers," *Journal of Human Resources*, 2009, *44.*

Evans, William N., Melinda S. Morrill, and Stephen T. Parente, "Measuring inappropriate medical diagnosis and treatment in survey data: The case of ADHD among school-age children," *Journal of Health Economics*, 2010, *29.*

Fan, Jianqing and Irene Gijbels, *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, Vol. 66, CRC Press, 1996.

Fiorini, Mario and Katrien Stevens, "Scrutinizing the Monotonicity Assumption in IV and fuzzy RD designs*," *Oxford Bulletin of Economics and Statistics*, 2021, *83.*

Fredriksson, Peter and Björn Öckert, "Life-cycle effects of age at school start," *Economic Journal*, 2014, *124.*

Gong, Jie, Yi Lu, and Hong Song, "The Effect of Teacher Gender on Students' Academic and Noncognitive Outcomes," *Journal of Labor Economics*, 2018, *36*, 743–778.

_ , _ , and _ , "Gender Peer Effects on Students' Academic and Noncognitive Outcomes Evidence and Mechanisms," *Journal of Human Resources*, 2021, *56* (3), 686–710.

Heckman, James J., Jora Stixrud, and Sergio Urzua, "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior," *Journal of Labor Economics*, 2006, *24*, 411–482.

Huang, Francis L., "Investigating the Prevalence of Academic Redshirting Using Population-Level Data," *AERA Open*, 2015, *1.*

Jackson, C. Kirabo, Shanette C. Porter, John Q. Easton, Alyssa Blanchard, and Sebastián Kiguel, "School Effects on Socioemotional Development, School-Based Arrests, and Educational Attainment," *American Economic Review: Insights*, 12 2020, *2*, 491–508.

Johansen, Eva Rye, "Relative age for grade and adolescent risky health behavior," *Journal of Health Economics*, 2021, *76.*

Kim, Taehoon, "The Effect of Age at School Entry on Long-term Educational Attainment in Korea," *Korean Journal of Labor Economics*, 2011, *34* (1), 1–32.

_ , "Age culture, school-entry cutoff, and the choices of birth month and school-entry timing in South Korea," *Journal of Demographic Economics*, 2021, *87.*

Kraft, Matthew A., "Teacher effects on complex cognitive skills and social-emotional competencies," *Journal of Human Resources*, 2019, *54*, 1–36.

Landersø, Rasmus, Helena Skyt Nielsen, and Marianne Simonsen, "School Starting Age and the Crime-age Profile," *Economic Journal*, 2017, *127.*

Landersø, Rasmus Kløve, Helena Skyt Nielsen, and Marianne Simonsen, "Effects of School Starting Age on the Family," *Journal of Human Resources*, 2020, *55.*

Lee, David S. and David Card, "Regression discontinuity inference with specification error," *Journal of Econometrics*, 2008, *142.*

Lindqvist, Erik and Roine Vestman, "The labor market returns to cognitive and noncognitive ability: Evidence from the swedish enlistment," *American Economic Journal: Applied Economics*, 2011, *3*, 101–128.

Lubotsky, Darren and Robert Kaestner, "Do 'Skills Beget Skills'? Evidence on the effect of kindergarten entrance age on the evolution of cognitive and non-cognitive skill gaps in childhood," *Economics of Education Review*, 2016, *53.*

McCrary, Justin, "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 2008, *142.*

Pei, Zhuan, David S. Lee, David Card, and Andrea Weber, "Local Polynomial Order in Regression Discontinuity Designs," *Journal of Business and Economic Statistics*, 2022, *40*, 1259–1267.

Pintrich, Paul R., David A. F. Smith, Teresa Garcia, and Wilbert J. Mckeachie, "Reliability and Predictive Validity of the Motivated Strategies for Learning Questionnaire (Mslq)," *Educational and Psychological Measurement*, 1993, *53.*

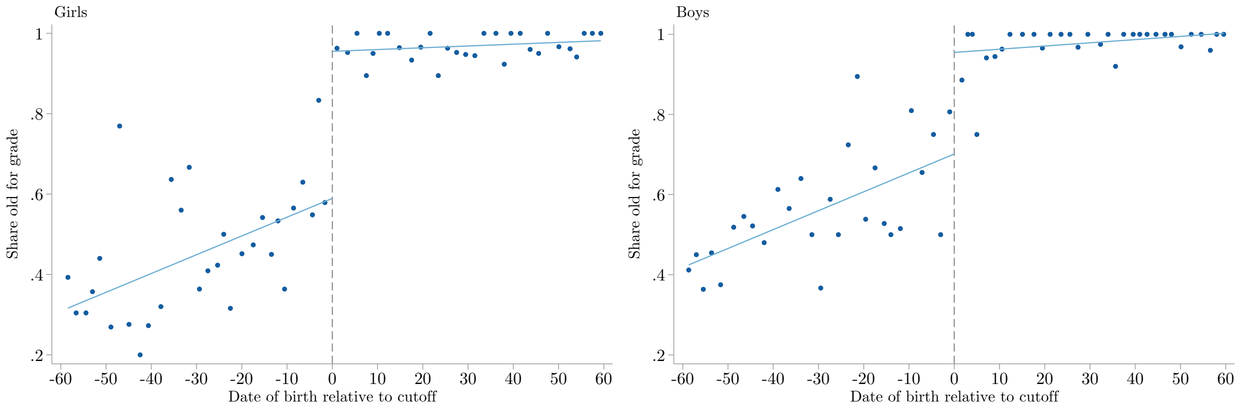
Shigeoka, Hitoshi, "SCHOOL ENTRY CUTOFF DATE AND THE TIMING OF BIRTHS," 2015.

Snyder, Thomas D and Sally A Dillow, "Digest of Education Statistics, 2012," Technical Report NCES 2014-015, National Center for Education Statistics, Washington, DC 2013.

Weinberger, Catherine J., "The Increasing Complementarity between Cognitive and Social Skills," *Review of Economics and Statistics*, 2014, *96.*

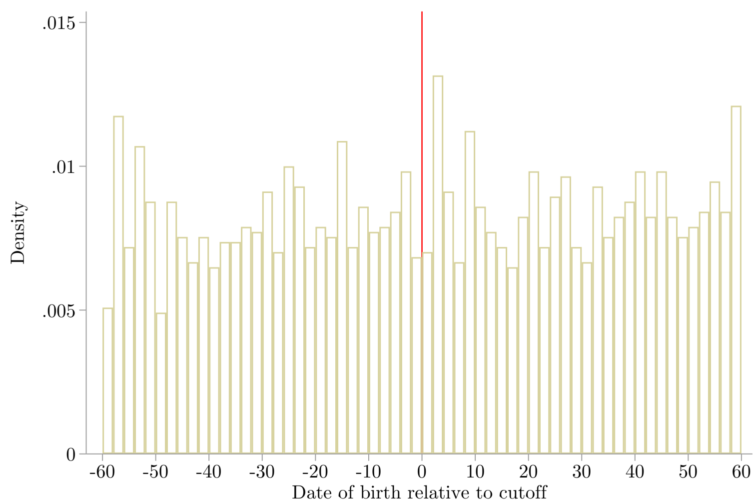
Figures

Figure 1: Share of Old-for-Grade Children among Girls (Left) and Boys (Right)



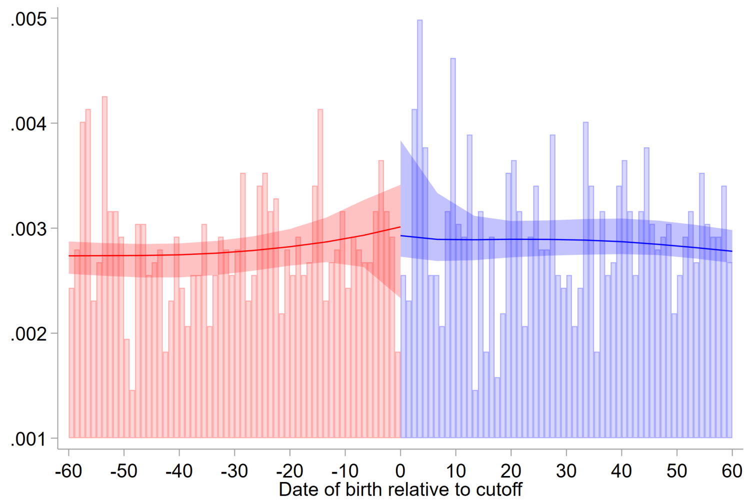
The share of old-for-grade children born between January and April is shown. The sample I use includes fourth grade and seventh grade cohorts in SELS 2010. The bin size is set to two days.

Figure 2: Density of Births Born Around the Cutoff



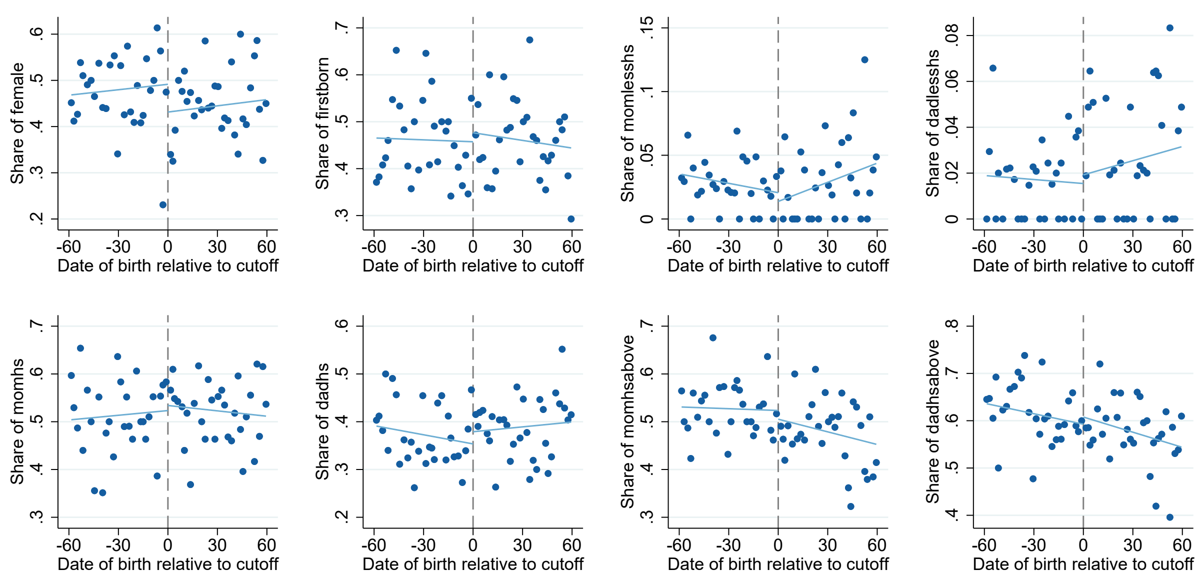
The density of births born between January and April is shown. The sample I use includes fourth grade and seventh grade cohorts in SELS 2010. The bin size is set to two days.

Figure 3: Manipulation Testing



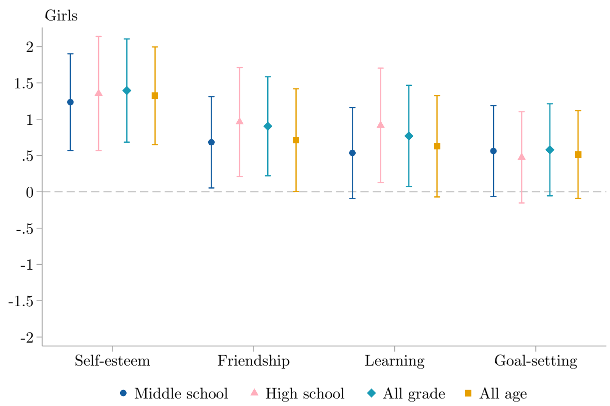
The figure shows manipulation testing around the cutoff following Cattaneo et al. (2020). I use 60-day bandwidths and fourth grade and seventh grade cohorts in SELS 2010.

Figure 4: Comparison of Predetermined Characteristics Before and After the Cutoff

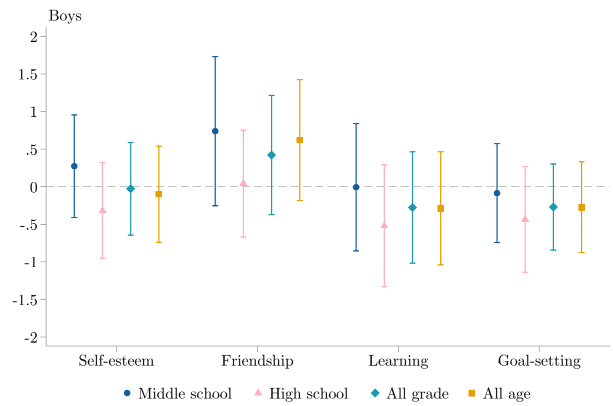


Individual predetermined characteristics include female, firstborn, and parents' educational attainment. Parents' educational attainment is divided into three categories: less than high school, high school graduates, and some college or more. I also include missing indicators for female, firstborn, and parents' educational attainment. I use 60-day bandwidths. The bin size is set to two days.

Figure 5: Effects of the School Starting Age on Girls' and Boys' Social-Emotional Skills



(a) Girls



(b) Boys

All regressions use local linear polynomials, a triangular kernel, and a 60-day bandwidth and include cohort fixed effects. Standard errors are clustered at the date of birth level. Middle school (MS) outcomes are averaged across seventh-ninth grades, high school (HS) outcomes are averaged across tenth-twelfth grades, and all grade outcomes are averaged across seventh-twelfth grades. All age outcomes are averaged across ages 13-17. All outcomes are standardized with a mean of zero and a unit variance.

Tables

Table 1: Summary Statistics of Outcome Indices by Cutoff and Gender

	Girls		Boys	
	Before cutoff	After cutoff	Before cutoff	After cutoff
Panel A: Self-esteem				
grades 7 to 9	3.597	3.699	3.728	3.718
	(0.707)	(0.735)	(0.683)	(0.666)
grades 10 to 12	3.623	3.667	3.807	3.778
	(0.705)	(0.694)	(0.665)	(0.701)
Panel B: Friendship				
grades 7 to 9	4.221	4.264	4.206	4.265
	(0.497)	(0.497)	(0.557)	(0.532)
grades 10 to 12	4.185	4.209	4.231	4.212
	(0.508)	(0.528)	(0.525)	(0.536)
Panel C: Learning approaches				
grades 7 to 9	3.275	3.342	3.333	3.339
	(0.635)	(0.622)	(0.650)	(0.622)
grades 10 to 12	3.314	3.351	3.336	3.315
	(0.633)	(0.605)	(0.605)	(0.634)
Panel D: Goal-setting				
grades 7 to 9	3.695	3.744	3.791	3.745
	(0.676)	(0.698)	(0.662)	(0.666)
grades 10 to 12	3.704	3.748	3.731	3.693
	(0.625)	(0.630)	(0.631)	(0.701)
Observations	648	648	700	809

Average and standard deviation before standardization to mean zero and a unit variance are shown. Standard deviations in parentheses. The sample I use includes fourth grade and seventh grade cohorts born between January and April. For each survey question, student response on 5 point scale: *1 'strongly disagree' 2 'disagree' 3 'neutral' 4 'agree' 5 'strongly agree'*, so the better social-emotional skill levels.

Table 2: First Stage Results

	Entire sample		Girls		Boys	
	(1)	(2)	(3)	(4)	(5)	(6)
Born after cutoff	0.323***	0.315***	0.376***	0.372***	0.278***	0.277***
	(0.026)	(0.026)	(0.034)	(0.033)	(0.048)	(0.048)
Observations	2825	2825	1287	1287	1494	1494
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls		Yes		Yes		Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are clustered at the date of birth level. Each cell reports results from a separate regression with a triangular kernel and 60-day bandwidths. All regressions include cohort fixed effects. The individual-level controls include a female indicator, a firstborn indicator, parents' educational attainment (categorized as less than high school, high school graduates, and some college or more), and indicators for missing values of these controls.

Table 3: Estimation Results using Regression Discontinuity Design: Middle School and High School

	Self-esteem		Friendship		Learning		Goal-setting	
	(1) MS	(2) HS	(3) MS	(4) HS	(5) MS	(6) HS	(7) MS	(8) HS
Panel A: Girls								
Old for grade	1.237***	1.356***	0.683**	0.964**	0.536*	0.916**	0.563*	0.475
	(0.340)	(0.401)	(0.321)	(0.383)	(0.320)	(0.403)	(0.320)	(0.321)
Observations	1216	1082	1216	1082	1214	1082	1217	1082
Panel B: Boys								
Old for grade	0.275	-0.320	0.743	0.042	-0.007	-0.523	-0.087	-0.438
	(0.349)	(0.326)	(0.510)	(0.364)	(0.435)	(0.415)	(0.338)	(0.360)
Observations	1416	1206	1417	1206	1417	1206	1417	1206
Panel C: Entire Sample								
Old for grade	0.823***	0.590*	0.667**	0.485*	0.239	0.164	0.251	-0.027
	(0.271)	(0.317)	(0.318)	(0.280)	(0.279)	(0.304)	(0.179)	(0.262)
Observations	2675	2324	2676	2324	2674	2324	2677	2324
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are clustered at the date of birth level. Each cell reports results from a separate regression with local linear polynomials, a triangular kernel, and a 60-day bandwidth. Middle school (MS) outcomes are averaged across grades 7 to 9, high school (HS) outcomes are averaged across grades 10 to 12. All outcomes are standardized with mean zero and unit variance. All regressions include cohort fixed effects.

Table 4: Estimation Results using Local Linear Regression Discontinuity Design: All Grade vs All Age

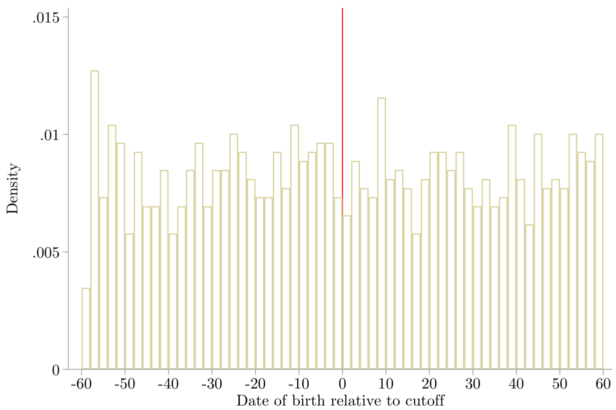
	Self-esteem		Friendship		Learning		Goal-setting	
	(1) Grade	(2) Age	(3) Grade	(4) Age	(5) Grade	(6) Age	(7) Grade	(8) Age
Panel A: Girls								
Old for grade	1.396***	1.324***	0.904***	0.712**	0.770**	0.628*	0.579*	0.515*
	(0.363)	(0.344)	(0.349)	(0.361)	(0.356)	(0.357)	(0.324)	(0.308)
Observations	1286	1274	1286	1274	1284	1273	1287	1274
Panel B: Boys								
Old for grade	-0.027	-0.099	0.423	0.624	-0.277	-0.289	-0.271	-0.274
	(0.316)	(0.328)	(0.407)	(0.413)	(0.379)	(0.385)	(0.293)	(0.310)
Observations	1493	1483	1494	1484	1494	1484	1494	1484
Panel C: Entire Sample								
Old for grade	0.762***	0.685**	0.635**	0.635**	0.217	0.131	0.130	0.0998
	(0.283)	(0.276)	(0.297)	(0.298)	(0.274)	(0.276)	(0.192)	(0.187)
Observations	2823	2801	2824	2802	2822	2801	2825	2802
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are clustered at the date of birth level. Each cell reports results from a separate regression with local linear polynomials, a triangular kernel, and a 60-day bandwidth. All grade outcomes are averaged across 7th-12th grades, and all age outcomes are averaged across age 13-17. All outcomes are standardized with mean zero and unit variance. All regressions include cohort fixed effects.

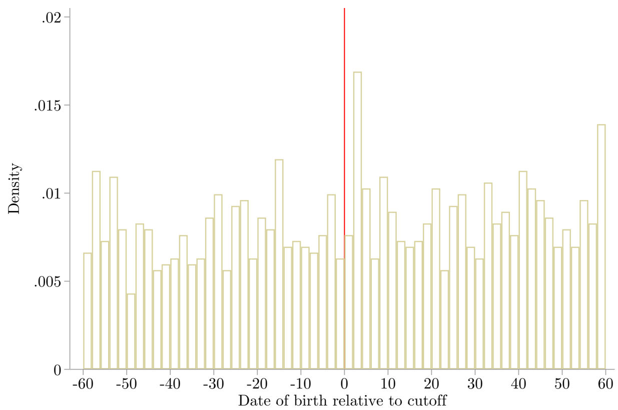
Appendix

Appendix Figures

Figure A1: Density of Births Born Around the Cutoff by Gender



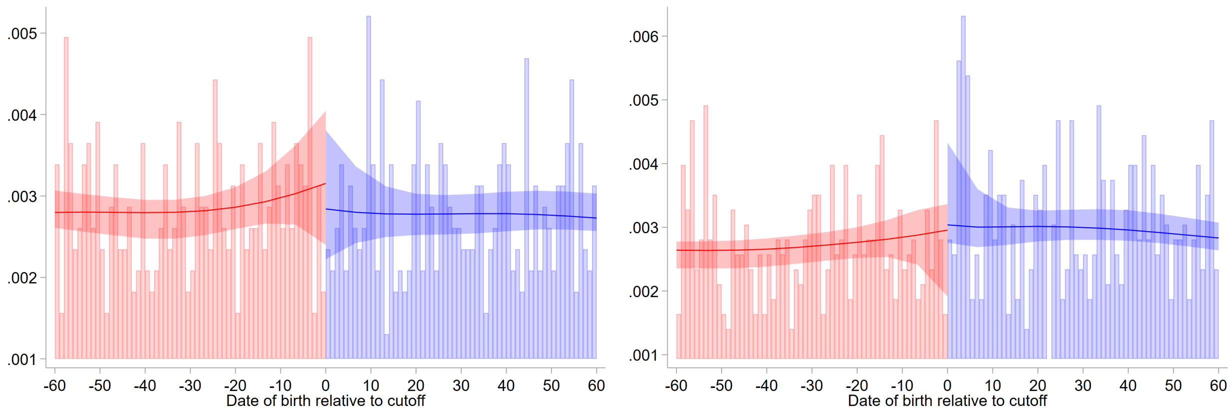
(a) Girls



(b) Boys

The figure shows the density of births born between January and April by gender. The sample I use includes girls and boys in fourth grade and seventh grade cohorts in SELS 2010. The bin size is set to two days.

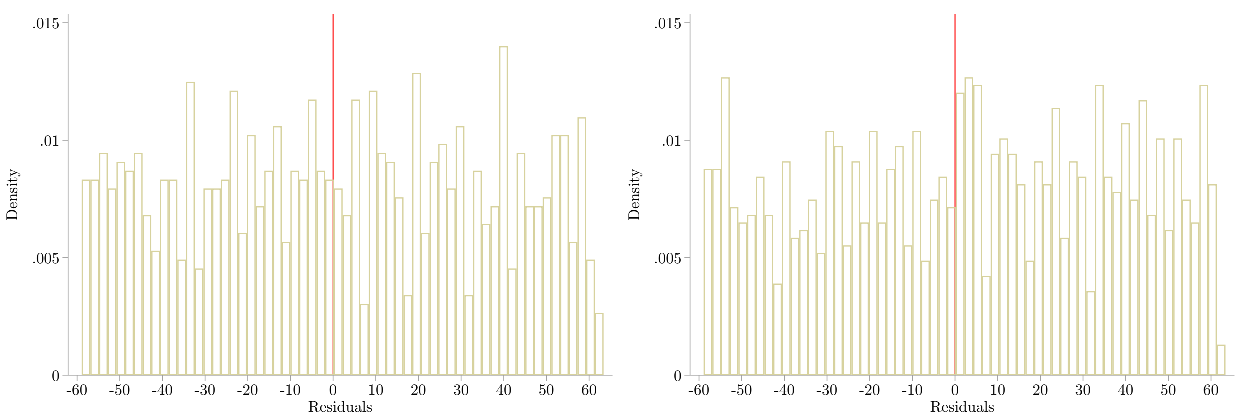
Figure A2: Manipulation Testing by Gender



(a) Girls (b) Boys

The figure shows the density of births born between January and April by gender. The sample I use includes girls and boys in fourth grade and seventh grade cohorts in SELS 2010. The bin size is set to two days.

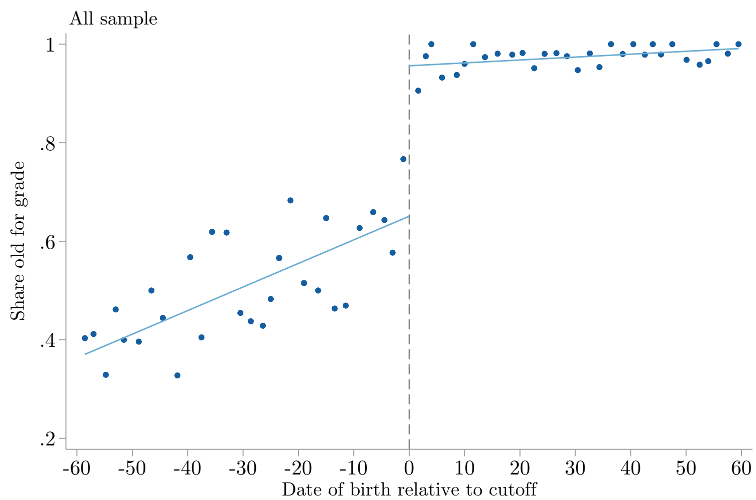
Figure A3: Density of Birth Born Around the Cutoff using Day-of-Week Fixed Effects



(a) Girls (b) Boys

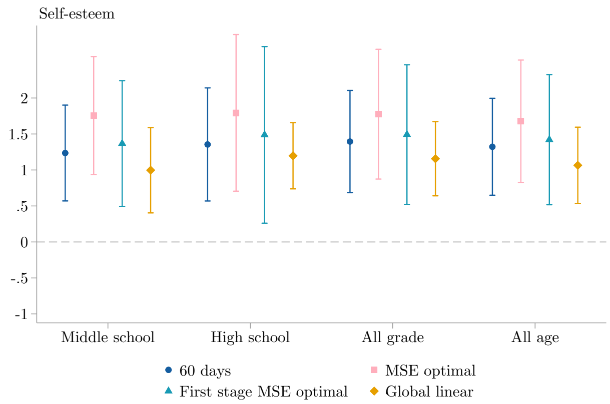
The figure shows the density of births born between January and April, using residuals from regressions incorporating day-of-week fixed effects and holidays Panel A shows the birth patterns of girls and Panel B shows the birth patterns of boys. The sample I use includes fourth grade and seventh grade cohorts in SELS 2010. The bin size is set to 2 days.

Figure A4: Share of Old for Grade Students

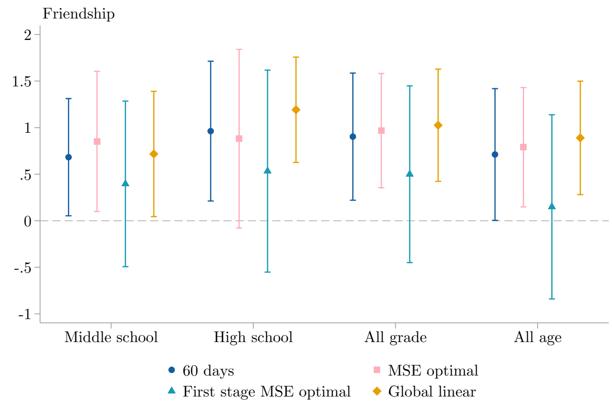


The share of old-for-grade children born between January and April is shown. The sample I use includes fourth grade and seventh grade cohorts in SELS 2010. The bin size is set to two days.

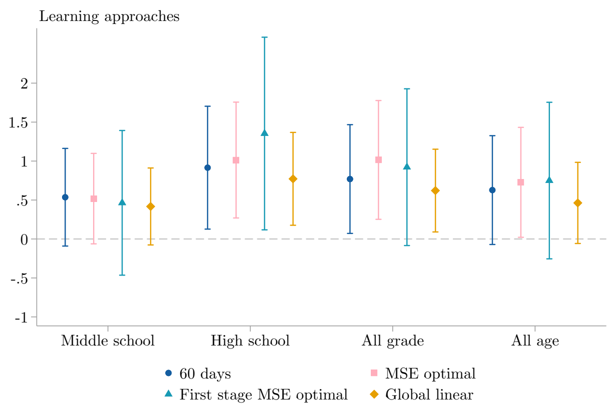
Figure A5: Sensitivity Check to Bandwidths (Girls)



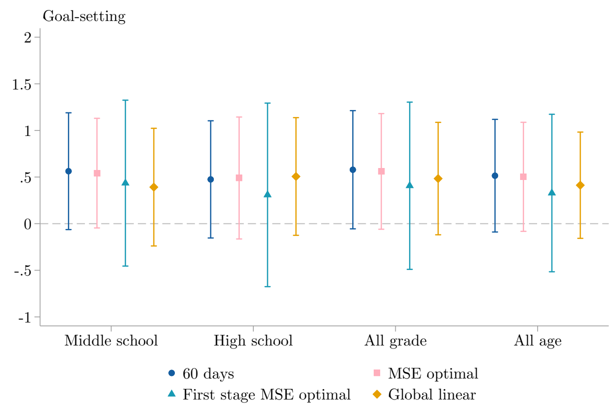
(a) Self-esteem



(b) Friendship



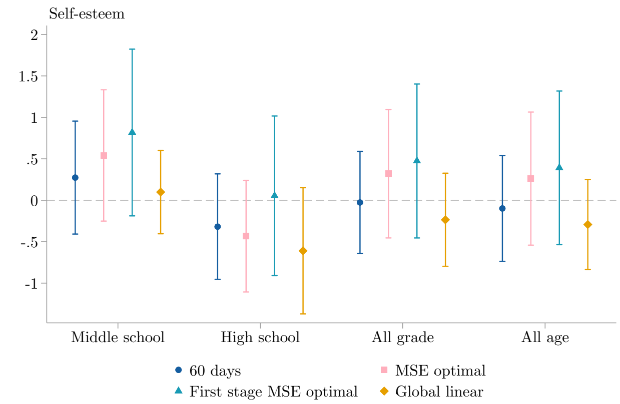
(c) Learning approach



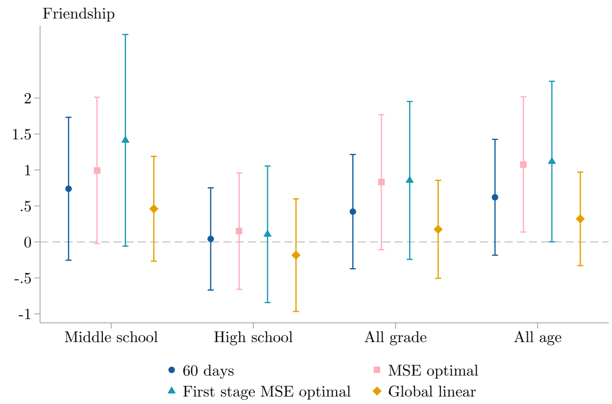
(d) Goal-setting

The estimates and corresponding 95 percent confidence intervals (CI) are shown. Circles represent estimates from the main regression using a 60-day bandwidth. Squares represent bias-corrected estimates and robust confidence intervals from the regression with mean squared error (MSE) optimal bandwidths following Calonico et al. (2014). Triangles represent bias-corrected estimates and robust confidence intervals from the regression with first-stage MSE optimal bandwidths. Lastly, diamonds represent estimates using global linear polynomials with a 60-day bandwidth. All estimates use a triangular kernel and include cohort fixed effects. Standard errors are clustered at the date of birth level. All outcomes are standardized with mean zero and unit variance.

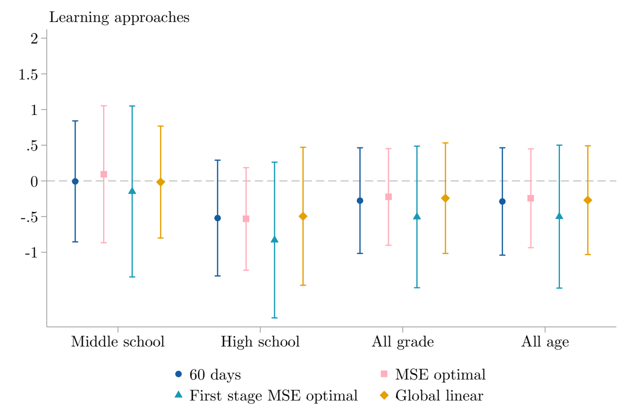
Figure A6: Sensitivity Check to Bandwidths (Boys)



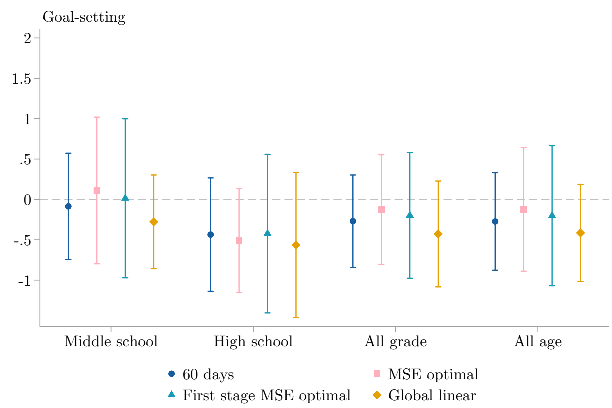
(a) Self-esteem



(b) Friendship



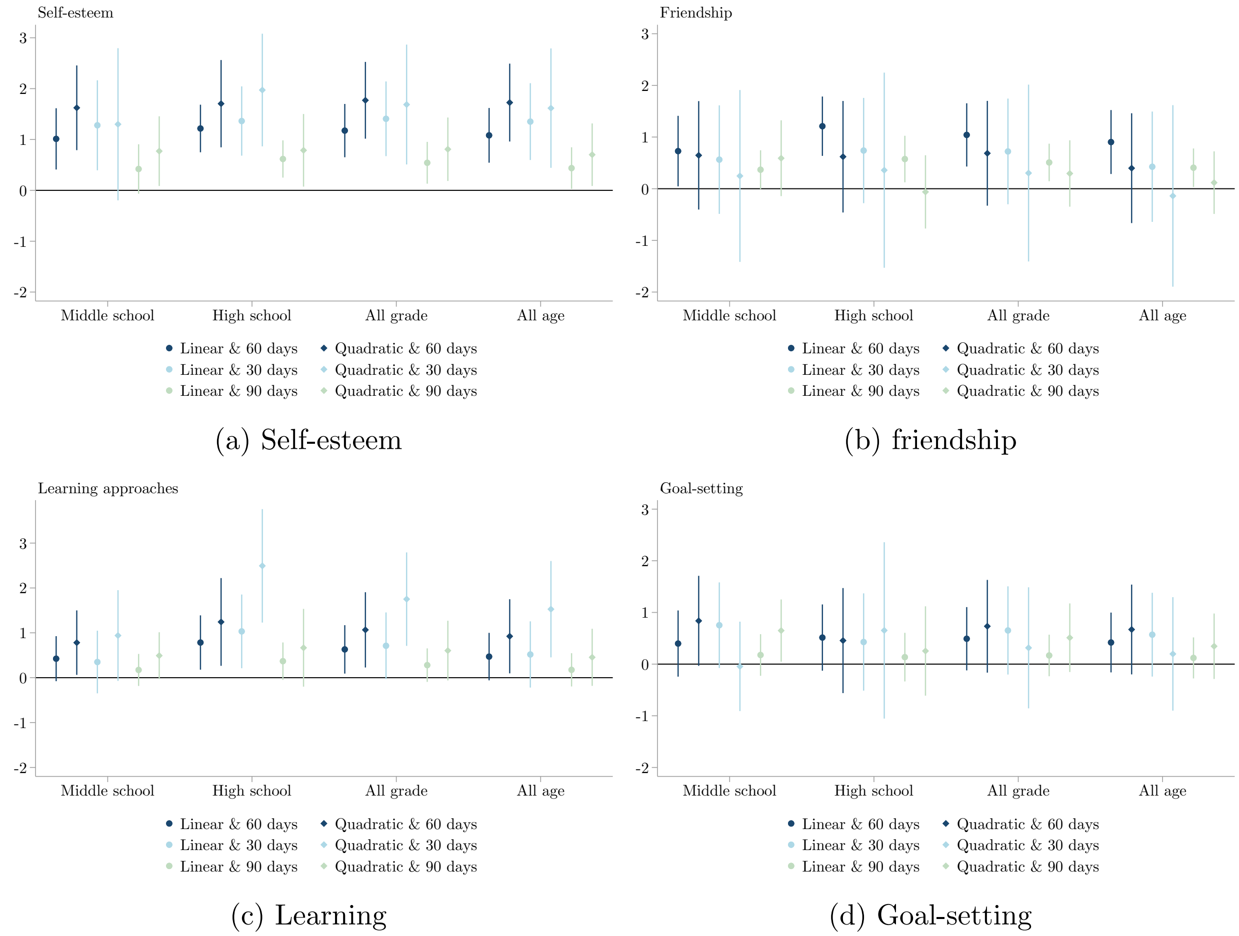
(c) Learning apprache



(d) Goal setting

The estimates and corresponding 95 percent confidence intervals (CI) are shown. Circles represent estimates from the main regression using a 60-day bandwidth. Squares represent bias-corrected estimates and robust confidence intervals from the regression with mean squared error (MSE) optimal bandwidths following Calonico et al. (2014). Triangles represent bias-corrected estimates and robust confidence intervals from the regression with first-stage MSE optimal bandwidths. Lastly, diamonds represent estimates using global linear polynomials with a 60-day bandwidth. All estimates use a triangular kernel and include cohort fixed effects. Standard errors are clustered at the date of birth level. All outcomes are standardized with mean zero and unit variance.

Figure A7: Sensitivity to Polynomials and Bandwidths (Girls)

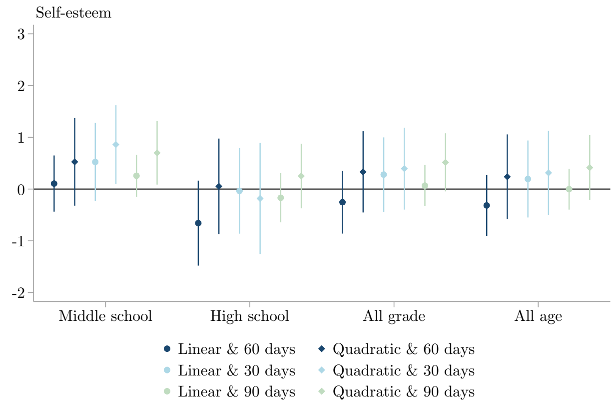


(a) Self-esteem (b) friendship

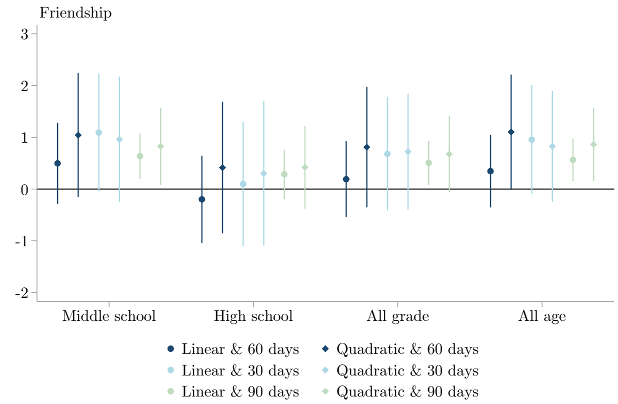
(c) Learning (d) Goal-setting

The estimates and corresponding 95 percent confidence intervals (CI) are shown. Circles represent the estimates using global linear polynomials. Diamonds represent the estimates using global quadratic polynomials. All estimates include cohort fixed effects. Standard errors are clustered at the date of birth level. All outcomes are standardized with mean zero and unit variance.

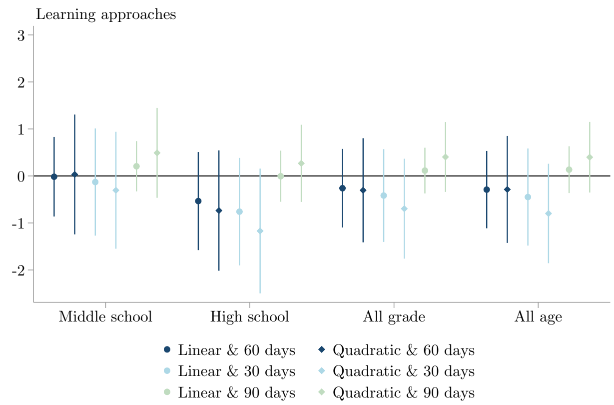
Figure A8: Sensitivity to Polynomials and Bandwidths (Boys)



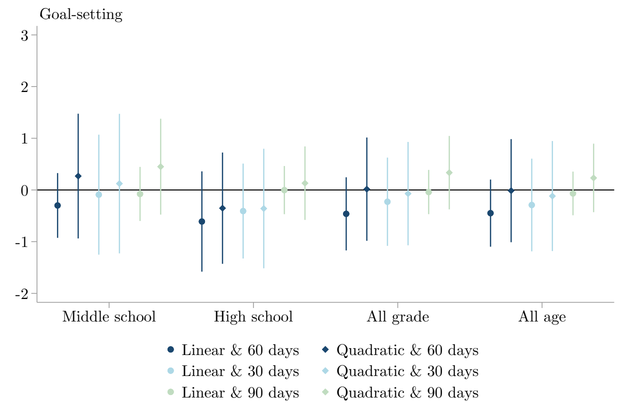
(a) Self-esteem



(b) friendship



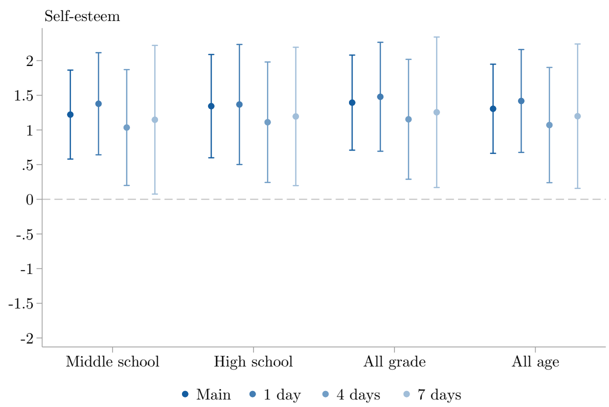
(c) Learning



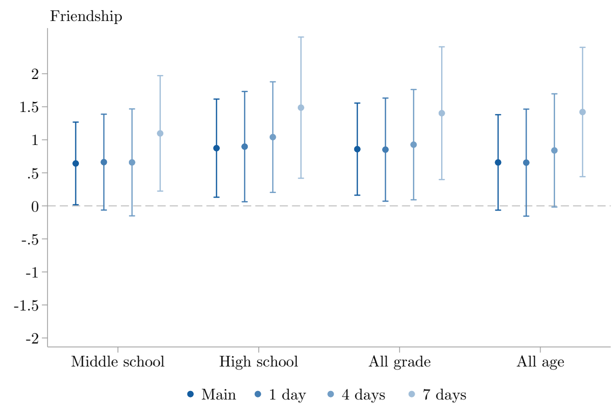
(d) Goal-setting

The estimates and corresponding 95 percent confidence intervals (CI) are shown. Circles represent the estimates using global linear polynomials. Diamonds represent the estimates using global quadratic polynomials. All estimates include cohort fixed effects. Standard errors are clustered at the date of birth level. All outcomes are standardized with mean zero and unit variance.

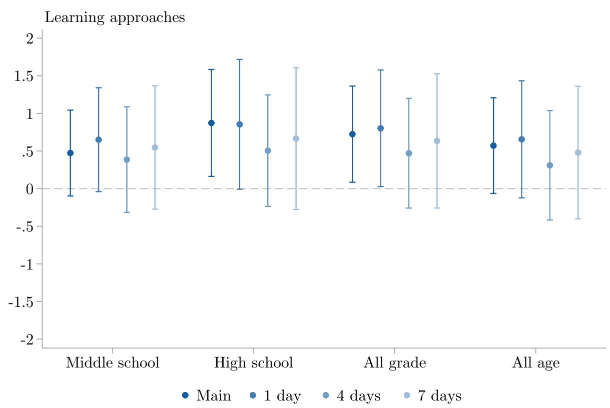
Figure A9: Estimation Results from Donut Regression Discontinuity Approaches (Girls)



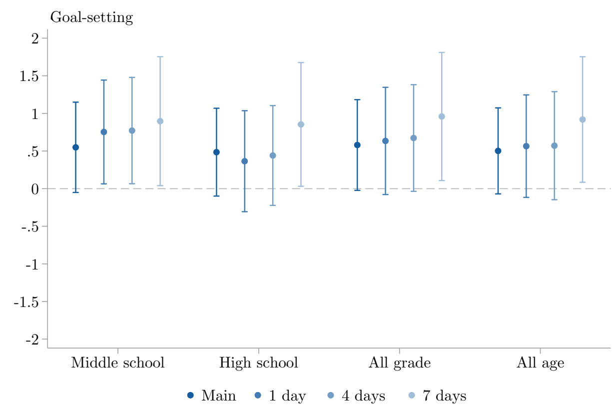
(a) Self-esteem



(b) Friendship



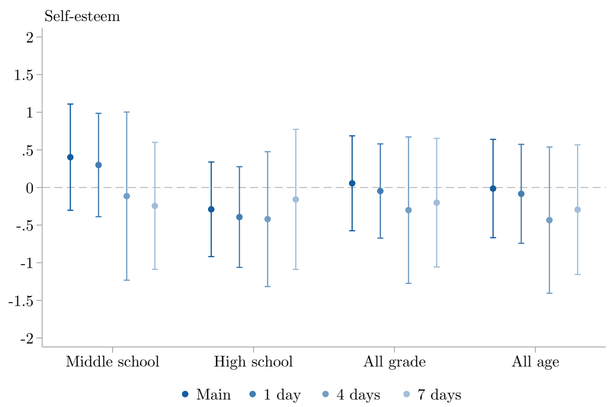
(c) Learning approach



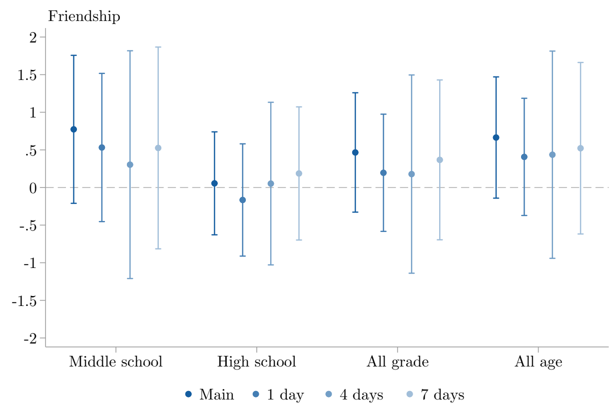
(d) Goal setting

The estimates and corresponding 95 percent confidence intervals (CI) are shown. All regressions use local linear polynomials, a triangular kernel, and a 60-day bandwidth and include cohort fixed effects. Standard errors are clustered at the date of birth level. All outcomes are standardized with mean zero and unit variance.

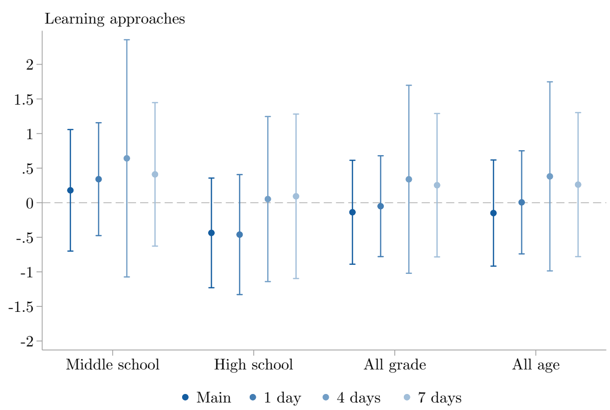
Figure A10: Estimation Results from Donut Regression Discontinuity Approaches (Boys)



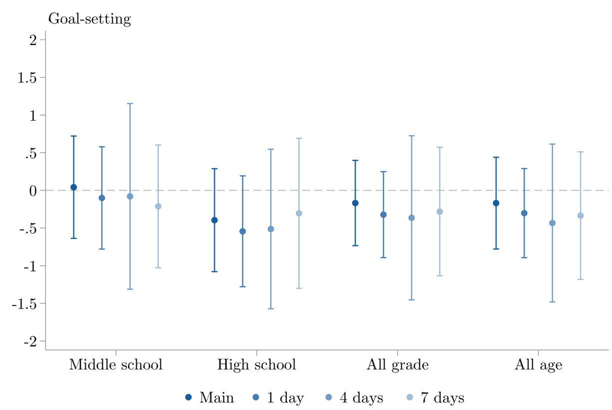
(a) Self-esteem



(b) Friendship



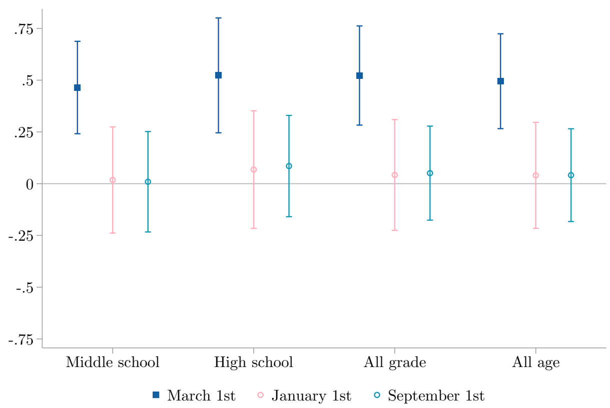
(c) Learning approach



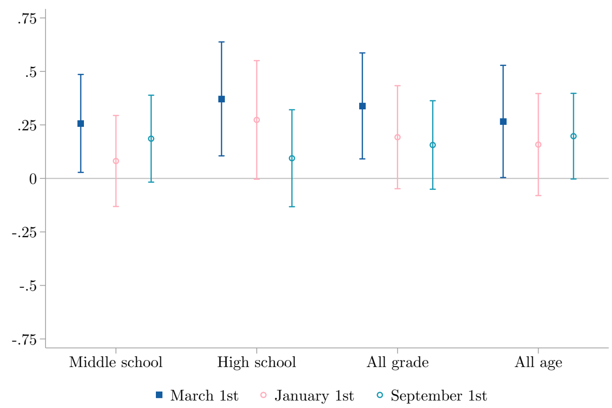
(d) Goal setting

The estimates and corresponding 95 percent confidence intervals (CI) are shown. All regressions use local linear polynomials, a triangular kernel, and a 60-day bandwidth and include cohort fixed effects. Standard errors are clustered at the date of birth level. All outcomes are standardized with mean zero and unit variance.

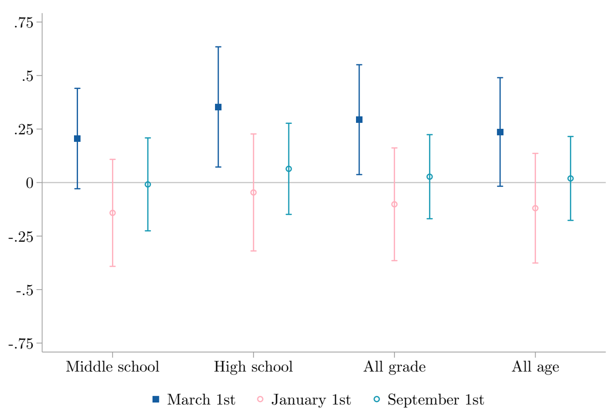
Figure A11: Comparison with Placebo Cutoffs (Girls)



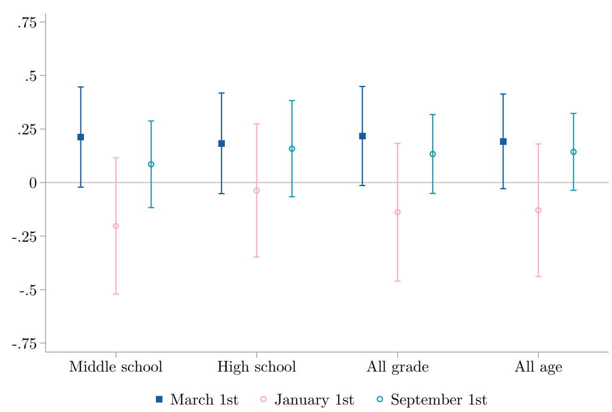
(a) Self-esteem



(b) Friendship



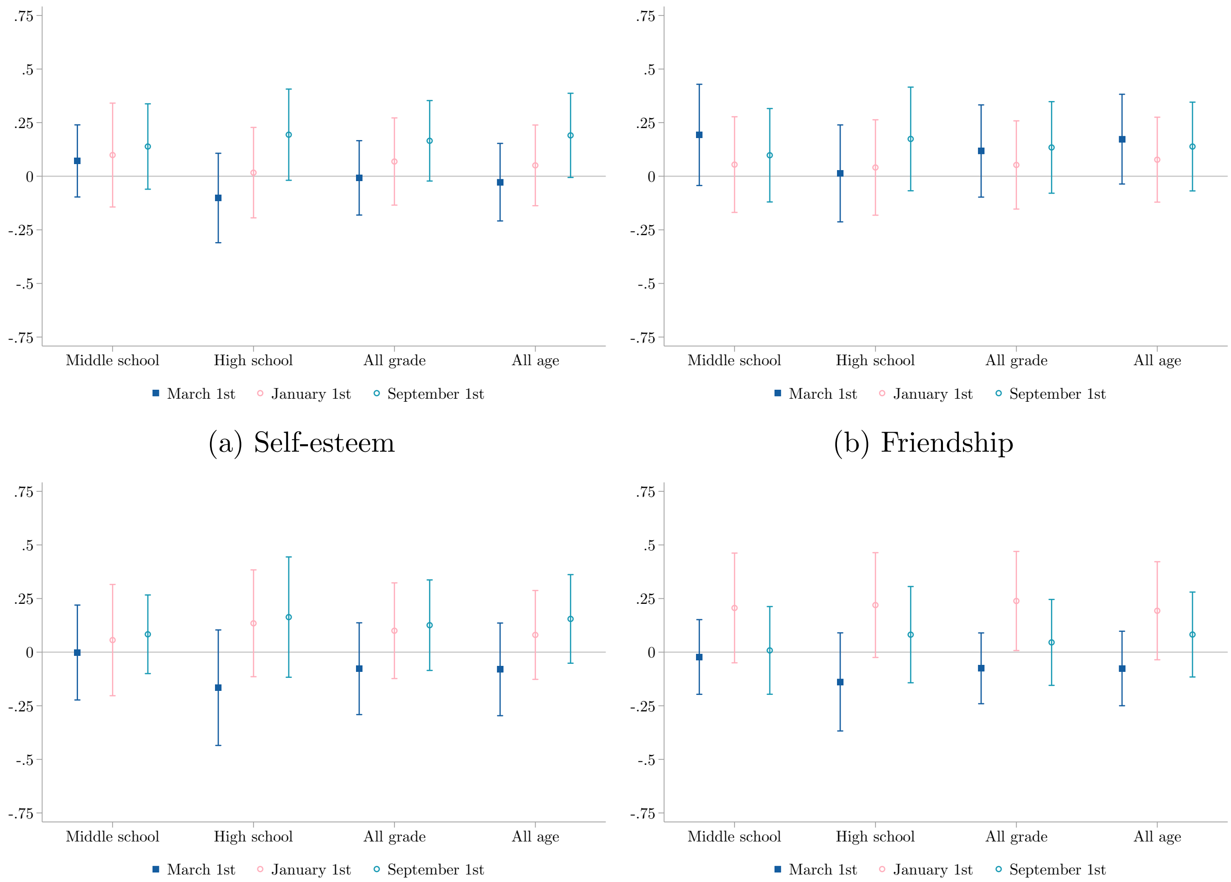
(c) Learning approach



(d) Goal-setting

The reduced-form estimates and corresponding 95 percent confidence intervals (CI) are shown. Squares represent estimates from the main specification. Hollow circles represent estimates using January 1st and September 1st as a placebo cutoff. All regressions use local linear polynomials, a triangular kernel, and a 60-day bandwidth and include cohort fixed effects. Standard errors are clustered at the date of birth level. All outcomes are standardized with mean zero and unit variance.

Figure A12: Comparison with Placebo Cutoffs (Boys)



(a) Self-esteem

(b) Friendship

(c) Learning appraches

(d) Goal setting

The reduced-form estimates and corresponding 95 percent confidence intervals (CI) are shown. Squares represent estimates from the main specification. Hollow circles represent estimates using January 1st and September 1st as a placebo cutoff. All regressions use local linear polynomials, a triangular kernel, and a 60-day bandwidth and include cohort fixed effects. Standard errors are clustered at the date of birth level. All outcomes are standardized with mean zero and unit variance.

Appendix Tables

Table A1: Average Outcome Indices by Cutoff and Gender: January - April Borns

	Girls		Boys	
	Before cutoff	After cutoff	Before cutoff	After cutoff
Panel A: Self-esteem				
grades 7 to 12	3.611	3.686	3.749	3.744
	(0.658)	(0.664)	(0.637)	(0.636)
ages 13 to 17	3.644	3.692	3.760	3.740
	(0.666)	(0.683)	(0.649)	(0.643)
Panel B: Friendship				
grades 7 to 12	4.201	4.236	4.217	4.236
	(0.463)	(0.479)	(0.500)	(0.499)
ages 13 to 17	4.211	4.244	4.226	4.251
	(0.476)	(0.479)	(0.505)	(0.499)
Panel C: Learning approaches				
grades 7 to 12	3.296	3.355	3.329	3.325
	(0.606)	(0.575)	(0.591)	(0.587)
ages 13 to 17	3.297	3.351	3.322	3.319
	(0.615)	(0.591)	(0.615)	(0.598)
Panel D: Goal-setting				
grades 7 to 12	3.714	3.753	3.759	3.729
	(0.603)	(0.632)	(0.587)	(0.630)
ages 13 to 17	3.706	3.743	3.739	3.723
	(0.617)	(0.647)	(0.616)	(0.636)
Observations	648	648	700	809

Average and standard deviation before standardization to mean zero and a unit variance are shown. Standard deviations in parentheses. The sample I use includes fourth grade and seventh grade cohorts born between January and April. For each survey question, student response on 5 point scale: *1 'strongly disagree' 2 'disagree' 3 'neutral' 4 'agree' 5 'strongly agree'*, so the better social-emotional skill levels.

Table A2: Root Mean Squared Error for Each Outcome: Middle School and High School

	Self-esteem		Friendship		Learning		Goal-setting	
	(1) MS	(2) HS	(3) MS	(4) HS	(5) MS	(6) HS	(7) MS	(8) HS
Panel A: Girls								
Linear	0.155	0.183	0.146	0.288	0.115	0.331	0.127	0.128
Quadratic	0.379	0.467	0.250	0.317	0.359	1.196	0.490	0.430
Cubic	1.095	1.229	1.170	0.934	1.734	2.299	1.017	1.228
Panel B: Boys								
Linear	0.425	0.277	0.583	0.168	0.215	0.217	0.229	0.168
Quadratic	0.503	0.405	0.510	0.321	0.783	0.961	0.443	0.441
Cubic	0.803	0.929	0.854	1.201	1.094	1.223	0.704	1.225

Each cell reports a result from a separate regression with a triangular kernel and 60-day bandwidth. Middle school (MS) outcomes are averaged across grades 7 to 9, high school (HS) outcomes are averaged across grades 10 to 12.

Table A3: Root Mean Squared Error for Each Outcome: All Grades and All Ages

	Self-esteem		Friendship		Learning		Goal-setting	
	(1) Grade	(2) Age	(3) Grade	(4) Age	(5) Grade	(6) Age	(7) Grade	(8) Age
Panel A: Girls								
Linear	0.159	0.158	0.236	0.359	0.158	0.147	0.129	0.127
Quadratic	0.371	0.350	0.249	0.262	0.636	0.547	0.317	0.312
Cubic	1.085	0.993	0.981	0.870	1.989	1.886	0.871	0.784
Panel B: Boys								
Linear	0.353	0.354	0.349	0.411	0.188	0.191	0.171	0.171
Quadratic	0.370	0.356	0.394	0.406	0.880	1.011	0.406	0.378
Cubic	0.827	0.796	0.852	0.848	1.198	1.355	0.895	0.841

Each cell reports a result from a separate regression with a triangular kernel and 60-day bandwidth. All grade outcomes are averaged across grades 7 to 12. All age outcomes are averaged across ages 13 to 17.

Table A4: First Stage Results using MSE Optimal Bandwidths

	All sample			Girls			Boys		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Born after cutoff	0.295***	0.283***	0.294***	0.359***	0.359***	0.357***	0.271***	0.270***	0.271***
	(0.033)	(0.034)	(0.033)	(0.043)	(0.041)	(0.044)	(0.073)	(0.072)	(0.072)
Bandwidth	31	32	31	38	38	37	40	40	40
Observations	1,505	1,548	1,505	834	834	813	1,014	1,014	1,014
Cohort FE	Yes	Yes		Yes	Yes		Yes	Yes	
Individual controls		Yes			Yes			Yes	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table presents bias-corrected estimates and robust standard errors using the mean squared error optimal bandwidths. Standard errors in parentheses are clustered at the date of birth level. Each cell reports results from a separate regression with a triangular kernel and linear polynomials. The individual-level controls include a female indicator, a firstborn indicator, parents' educational attainment (categorized as less than high school, high school graduates, and some college or more), and indicators for missing values of these controls.

Table A5: Estimation Results using MSE Optimal Bandwidths: Middle School and High school

	Self-esteem		Friendship		Learning		Goal-setting	
	(1) MS	(2) HS	(3) MS	(4) HS	(5) MS	(6) HS	(7) MS	(8) HS
Panel A: Girls								
Old for grade	1.296***	1.390***	0.657**	0.826**	0.429*	0.738**	0.444*	0.441
	(0.361)	(0.456)	(0.334)	(0.407)	(0.256)	(0.331)	(0.263)	(0.292)
Bandwidth	51	48	54	49	73	70	72	66
Observations	1035	870	1113	886	1487	1273	1478	1203
Panel B: Boys								
Old for grade	0.320	-0.402	0.672	0.036	-0.002	-0.439	-0.015	-0.452
	(0.359)	(0.296)	(0.467)	(0.362)	(0.424)	(0.326)	(0.396)	(0.295)
Bandwidth	56	69	66	60	62	78	42	77
Observations	1322	1396	1549	1224	1471	1546	1000	1538
Panel C: Entire Sample								
Old for grade	0.968***	0.645*	0.707**	0.388*	0.244	0.122	0.231	-0.046
	(0.318)	(0.366)	(0.336)	(0.230)	(0.281)	(0.238)	(0.192)	(0.215)
Bandwidth	42	46	55	72	61	74	42	73
Observations	1891	1800	2475	2789	2752	2862	1893	2829
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are clustered at date of birth level. Each cell reports results from a separate regression with a triangular kernel. All regressions include cohort fixed effects. Middle school (MS) outcomes are calculated as the average across 7th to 9th grades, while high school (HS) outcomes are calculated as the average across 10th to 12th grades. All outcomes are standardized with a mean of zero and a unit variance.

Table A6: Estimation Results using MSE Optimal Bandwidths: All Grade vs All Age

	Self-esteem		Friendship		Learning		Goal-setting	
	(1) Grade	(2) Age	(3) Grade	(4) Age	(5) Grade	(6) Age	(7) Grade	(8) Age
Panel A: Girls								
Old for grade	1.435***	1.368***	0.747***	0.624**	0.754**	0.591*	0.485*	0.425
	(0.395)	(0.373)	(0.272)	(0.281)	(0.348)	(0.340)	(0.278)	(0.262)
Bandwidth	47	47	75	75	62	63	70	70
Observations	1016	1007	1628	1616	1353	1354	1521	1508
Panel B: Boys								
Old for grade	0.113	0.077	0.530	0.729*	-0.229	-0.245	-0.207	-0.211
	(0.339)	(0.358)	(0.427)	(0.429)	(0.309)	(0.317)	(0.307)	(0.340)
Bandwidth	50	47	52	53	75	75	46	44
Observations	1246	1178	1296	1306	1825	1814	1161	1102
Panel C: All Sample								
Old for grade	0.880***	0.799**	0.574**	0.579**	0.198	0.128	0.138	0.102
	(0.338)	(0.323)	(0.264)	(0.266)	(0.232)	(0.250)	(0.200)	(0.198)
Bandwidth	41	41	68	68	71	68	51	49
Observations	1965	1952	3210	3188	3324	3187	2407	2304
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are clustered at date of birth level. Each cell reports results from a separate regression with a triangular kernel. All regressions include cohort fixed effects. All grade outcomes represent the average across 7th to 12th grades. All age outcomes are averaged across the ages of 13 to 17. All outcomes are standardized with a mean of zero and a unit variance.

Table A7: Estimation Results in Reduced-Form: Middle School and High School

	Self-esteem		Friendship		Learning		Goal-setting	
	(1) MS	(2) HS	(3) MS	(4) HS	(5) MS	(6) HS	(7) MS	(8) HS
Panel A: Girls								
Born after cutoff	0.464***	0.523***	0.257**	0.372***	0.206*	0.353**	0.213*	0.183
	(0.114)	(0.142)	(0.117)	(0.136)	(0.120)	(0.143)	(0.119)	(0.120)
Observations	1,216	1,082	1,216	1,082	1,214	1,082	1,217	1,082
Panel B: Boys								
Born after cutoff	0.072	-0.101	0.193	0.013	-0.002	-0.166	-0.022	-0.139
	(0.086)	(0.106)	(0.120)	(0.115)	(0.113)	(0.137)	(0.089)	(0.117)
Observations	1,416	1,206	1,417	1,206	1,417	1,206	1,417	1,206
Panel C: Entire Sample								
Born after cutoff	0.261***	0.201*	0.214**	0.169*	0.079	0.055	0.083	-0.011
	(0.077)	(0.105)	(0.093)	(0.095)	(0.086)	(0.103)	(0.055)	(0.090)
Observations	2,675	2,324	2,676	2,324	2,674	2,324	2,677	2,324
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are clustered at the date of birth level. Each cell reports results from a separate regression with local linear polynomials, a triangular kernel, and a 60-day bandwidth. Middle school (MS) outcomes are averaged across grades 7 to 9, high school (HS) outcomes are averaged across grades 10 to 12. All outcomes are standardized with mean zero and unit variance. All regressions include cohort fixed effects.

Table A8: Fix Grade vs Fix Age: Self-esteem

	(1) 7th	(2) age 13	(3) 8th	(4) age 14	(5) 9th	(6) age 15	(7) 10th	(8) age 16	(9) 11th	(10) age 17	(11) 12th
Panel A: Girls											
Old-for-grade	1.074***	0.957***	0.965***	0.844**	0.955***	1.006***	1.259***	1.399***	1.386***	1.273***	0.977**
	(0.300)	(0.301)	(0.320)	(0.344)	(0.360)	(0.356)	(0.395)	(0.405)	(0.387)	(0.418)	(0.411)
Observations	1206	1190	1157	1152	1128	1116	1072	1055	1031	1022	1010
Panel B: Boys											
Old-for-grade	0.239	-0.0925	0.044	0.166	0.373	-0.155	-0.145	0.057	-0.468	-0.398	-0.207
	(0.426)	(0.456)	(0.323)	(0.277)	(0.408)	(0.358)	(0.339)	(0.394)	(0.408)	(0.388)	(0.344)
Observations	1392	1381	1328	1325	1296	1276	1192	1180	1150	1143	1106
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are clustered at the date of birth level. Each cell reports results from a separate regression with local linear polynomials, a triangular kernel, and a 60-day bandwidth. All outcomes are standardized with mean zero and unit variance. All regressions include cohort fixed effects.

Table A9: Fix Grade vs Fix Age: Friendship

	(1) 7th	(2) age 13	(3) 8th	(4) age 14	(5) 9th	(6) age 15	(7) 10th	(8) age 16	(9) 11th	(10) age 17	(11) 12th
Panel A: Girls											
Old-for-grade	0.580	0.503	0.452*	0.396	0.622**	0.584*	0.691*	0.679	0.487	0.693*	1.318***
	(0.357)	(0.357)	(0.256)	(0.281)	(0.312)	(0.329)	(0.399)	(0.414)	(0.456)	(0.394)	(0.344)
Observations	1211	1195	1160	1153	1125	1114	1070	1052	1030	1024	1014
Panel B: Boys											
Old-for-grade	0.199	0.845*	0.866*	1.067**	0.688	0.239	0.0245	-0.251	0.284	0.498	-0.169
	(0.432)	(0.459)	(0.502)	(0.496)	(0.533)	(0.411)	(0.406)	(0.369)	(0.468)	(0.411)	(0.402)
Observations	1405	1390	1335	1330	1299	1278	1185	1177	1150	1143	1108
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are clustered at the date of birth level. Each cell reports results from a separate regression with local linear polynomials, a triangular kernel, and a 60-day bandwidth. All outcomes are standardized with mean zero and unit variance. All regressions include cohort fixed effects.

Table A10: Fix Grade vs Fix Age: Learning Approaches

	(1) 7th	(2) age 13	(3) 8th	(4) age 14	(5) 9th	(6) age 15	(7) 10th	(8) age 16	(9) 11th	(10) age 17	(11) 12th
Panel A: Girls											
Old-for-grade	0.553**	0.544*	0.481	0.290	0.339	0.323	0.765**	0.828**	0.837**	0.719*	0.905**
	(0.270)	(0.328)	(0.326)	(0.325)	(0.348)	(0.313)	(0.344)	(0.402)	(0.426)	(0.436)	(0.458)
Observations	1199	1183	1157	1154	1129	1112	1065	1051	1032	1022	1009
Panel B: Boys											
Old-for-grade	0.317	0.301	-0.022	0.100	-0.349	-0.699**	-0.456	-0.253	-0.825**	-0.678	-0.094
	(0.467)	(0.466)	(0.454)	(0.453)	(0.384)	(0.336)	(0.426)	(0.454)	(0.420)	(0.481)	(0.457)
Observations	1387	1375	1325	1322	1291	1270	1186	1177	1147	1140	1106
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are clustered at the date of birth level. Each cell reports results from a separate regression with local linear polynomials, a triangular kernel, and a 60-day bandwidth. All outcomes are standardized with mean zero and unit variance. All regressions include cohort fixed effects.

Table A11: Fix Grade vs Fix Age: Goal-setting Mindsets

	(1) 7th	(2) age 13	(3) 8th	(4) age 14	(5) 9th	(6) age 15	(7) 10th	(8) age 16	(9) 11th	(10) age 17	(11) 12th
Panel A: Girls											
Old-for-grade	0.399	0.449	0.507*	0.297	0.474	0.454	0.591**	0.587*	0.216	0.228	0.405
	(0.297)	(0.318)	(0.276)	(0.284)	(0.370)	(0.355)	(0.265)	(0.301)	(0.365)	(0.375)	(0.394)
Observations	1199	1181	1154	1152	1128	1115	1070	1053	1033	1024	1010
Panel B: Boys											
Old-for-grade	0.005	-0.288	-0.370	0.093	0.137	-0.454	-0.606	-0.325	-0.323	-0.113	-0.129
	(0.385)	(0.343)	(0.301)	(0.294)	(0.365)	(0.357)	(0.386)	(0.422)	(0.378)	(0.379)	(0.343)
Observations	1397	1385	1327	1323	1298	1274	1190	1182	1150	1143	1108
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are clustered at the date of birth level. Each cell reports results from a separate regression with local linear polynomials, a triangular kernel, and a 60-day bandwidth. All outcomes are standardized with mean zero and unit variance. All regressions include cohort fixed effects.

Table A12: Difference in Height and Weight

	Fourth grade cohort		Seventh grade cohort	
	(1) Height(cm)	(2) Weight(kg)	(3) Height(cm)	(4) Weight(kg)
Panel A: Girls				
Old for grade	5.878	-5.418	1.235	3.220
	(4.507)	(7.120)	(1.625)	(2.715)
Baseline mean	137.318	33.866	156.441	46.101
Observations	575	575	622	578
Panel B: Boys				
Old for grade	3.445	3.252	-1.059	1.634
	(3.356)	(4.151)	(5.681)	(11.608)
Baseline mean	137.196	35.455	156.936	49.096
Observations	605	605	778	758

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are clustered at the date of birth level. Each cell reports results from a separate regression with local linear polynomials, a triangular kernel, and a 60-day bandwidth. 2.54cm = 1in & 1kg = 2.205lb

Table A13: Sensitivity Check: Self-esteem in Middle School and High School

	Middle school				High school			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Girls								
Old for grade	1.236***	1.180***	1.229***	1.276***	1.355***	1.306***	1.311***	1.411***
	(0.340)	(0.327)	(0.334)	(0.338)	(0.401)	(0.382)	(0.386)	(0.386)
Observations	1216	1216	1216	1216	1082	1082	1082	1082
Panel B: Boys								
Old for grade	0.273	0.401	0.271	0.235	-0.319	-0.250	-0.315	-0.296
	(0.348)	(0.355)	(0.344)	(0.314)	(0.325)	(0.328)	(0.321)	(0.309)
Observations	1416	1416	1416	1416	1206	1206	1206	1206
Cohort FE	Yes	Yes		Yes	Yes	Yes		Yes
Individual controls		Yes				Yes		
Day-of-week FE				Yes				Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are clustered at the date of birth level. Each cell reports results from a separate regression with local linear polynomials, a triangular kernel, and a 60-day bandwidth. Middle school outcomes are averaged across 7th-9th grades, and high school outcomes are averaged across 10th-12th grades. All outcomes are standardized with mean zero and unit variance.

Table A14: Sensitivity Check: Friendship in Middle School and High School

	Middle school				High school			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Girls								
Old for grade	0.683**	0.624*	0.686**	0.718**	0.963**	0.858**	0.967**	1.017***
	(0.321)	(0.319)	(0.316)	(0.323)	(0.383)	(0.380)	(0.377)	(0.386)
Observations	1216	1216	1216	1216	1082	1082	1082	1082
Panel B: Boys								
Old for grade	0.739	0.768	0.746	0.702	0.042	0.068	0.031	0.035
	(0.507)	(0.495)	(0.504)	(0.499)	(0.362)	(0.345)	(0.357)	(0.352)
Observations	1417	1417	1417	1417	1206	1206	1206	1206
Cohort FE	Yes	Yes		Yes	Yes	Yes		Yes
Individual controls		Yes				Yes		
Day-of-week FE				Yes				Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are clustered at the date of birth level. Each cell reports results from a separate regression with local linear polynomials, a triangular kernel, and a 60-day bandwidth. Middle school outcomes are averaged across 7th-9th grades, and high school outcomes are averaged across 10th-12th grades. All outcomes are standardized with mean zero and unit variance. All regressions include cohort fixed effects.

Table A15: Sensitivity Check: Learning Approaches in Middle School and High School

	Middle school				High school			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Girls								
Old for grade	0.536*	0.431	0.551*	0.557*	0.915**	0.829**	0.899**	0.964**
	(0.320)	(0.284)	(0.316)	(0.312)	(0.402)	(0.356)	(0.393)	(0.396)
Observations	1214	1214	1214	1214	1082	1082	1082	1082
Panel B: Boys								
Old for grade	-0.007	0.181	-0.008	-0.033	-0.521	-0.407	-0.524	-0.504
	(0.432)	(0.436)	(0.428)	(0.419)	(0.414)	(0.397)	(0.409)	(0.377)
Observations	1417	1417	1417	1417	1206	1206	1206	1206
Cohort FE	Yes	Yes		Yes	Yes	Yes		Yes
Individual controls		Yes				Yes		
Day-of-week FE				Yes				Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are clustered at the date of birth level. Each cell reports results from a separate regression with local linear polynomials, a triangular kernel, and a 60-day bandwidth. Middle school outcomes are averaged across 7th-9th grades, and high school outcomes are averaged across 10th-12th grades. All outcomes are standardized with mean zero and unit variance. All regressions include cohort fixed effects.

Table A16: Sensitivity Check: Goal-setting in Middle School and High School

	Middle school				High school			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Girls								
Old for grade	0.563*	0.510*	0.562*	0.581*	0.475	0.459	0.438	0.504
	(0.320)	(0.309)	(0.315)	(0.322)	(0.321)	(0.299)	(0.305)	(0.321)
Observations	1217	1217	1217	1217	1082	1082	1082	1082
Panel B: Boys								
Old for grade	-0.086	0.047	-0.069	-0.082	-0.436	-0.383	-0.421	-0.397
	(0.336)	(0.346)	(0.335)	(0.331)	(0.358)	(0.346)	(0.352)	(0.310)
Observations	1417	1417	1417	1417	1206	1206	1206	1206
Cohort FE	Yes	Yes		Yes	Yes	Yes		Yes
Individual controls		Yes				Yes		
Day-of-week FE				Yes				Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are clustered at the date of birth level. Each cell reports results from a separate regression with local linear polynomials, a triangular kernel, and a 60-day bandwidth. Middle school outcomes are averaged across 7th-9th grades, and high school outcomes are averaged across 10th-12th grades. All outcomes are standardized with mean zero and unit variance. All regressions include cohort fixed effects.

Table A17: Robust to Inclusion of School Fixed Effects: Self-esteem

	Middle school			High school			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Girls							
Old for grade	1.237***	1.237***	1.105***	1.356***	1.357***	0.971***	1.300***
	(0.340)	(0.336)	(0.219)	(0.401)	(0.388)	(0.223)	(0.263)
Observations	1,216	1,216	1,216	1,082	1,012	1,012	1,077
Panel B: Boys							
Old for grade	0.275	0.275	0.474	-0.320	-0.175	-0.109	-0.531*
	(0.349)	(0.422)	(0.340)	(0.326)	(0.434)	(0.319)	(0.295)
Observations	1,416	1,416	1,416	1,206	1,129	1,129	1,204
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Middle school FE			Yes			Yes	
High school FE							Yes
Cluster	DOB	MS	MS	DOB	MS	MS	HS

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses are clustered at the date of birth level (DOB), middle school level (MS), and high school level (HS). Each cell reports results from a separate regression with local linear polynomials, a triangular kernel, and a 60-day bandwidth. Middle school outcomes are averaged across 7th-9th grades, and high school outcomes are averaged across 10th-12th grades. All outcomes are standardized with mean zero and unit variance. All regressions include cohort fixed effects.

Table A18: Reliability Measures: Cronbach's α

	Middle school	High school	
		Academic	Vocational
Self-esteem	0.941	0.911	0.907
Friendship	0.782	0.785	0.787
Learning method	0.835	0.760	0.723
Learning effort	0.823	0.767	0.721
Learning attitude	0.776	0.717	0.655
Goal-setting	0.902	0.861	0.858

Cronbach's α is measured in the fourth survey wave.