

Month-of-Birth Effects on Skills and Skill Formation

By

Shintaro Yamaguchi (The University of Tokyo)

Hirotake Ito (Keio University)

Makiko Nakamuro (Keio University)

July 2020

CREPE DISCUSSION PAPER NO. 76



CENTER FOR RESEARCH AND EDUCATION FOR POLICY EVALUATION (CREPE)

THE UNIVERSITY OF TOKYO

<http://www.crepe.e.u-tokyo.ac.jp/>

Month-of-Birth Effects on Skills and Skill Formation*

Shintaro Yamaguchi[†] Hirotake Ito[‡] Makiko Nakamuro[§]

July 11, 2020

Abstract

We estimate month-of-birth effects on cognitive and noncognitive skills, as well as factors relevant to skill formation. Our estimates indicate that younger students in a given grade cohort have lower cognitive and noncognitive skills. To shed light on the underlying mechanisms, we also examine children's time use and interpersonal relationships. We find evidence of compensatory investment in cognitive skills by younger children, however, they are also more likely to have poor-quality relationships with their classmates and teachers, which may harm the long-term development of noncognitive skills. Our findings on skill formation have implications for efficiency and equity.

Keywords: relative age effect, cognitive skill, non-cognitive skill

JEL Codes: I210, I240, J24

*We thank an anonymous province for providing data and research fund. We gratefully acknowledge financial support from MEXT KAKENHI (16H02022) and JSPS KAKENHI (20H01510) and administrative support for Research Institute of Economy, Trade and Industry. We thank seminar participants at various universities and conferences for their comments and especially Daiji Kawaguchi and Michihito Ando for their invaluable comments on our draft.

[†]Corresponding author. Faculty of Economics, the University of Tokyo. Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan. Email: syamaguchi@e.u-tokyo.ac.jp.

[‡]Graduate School of Media and Governance, Keio University, 5322 Endo, Fujisawa-shi, Kanagawa 252-0882 Japan, Email: itouhrtk@keio.ac.jp.

[§]Graduate School of Media and Governance, Keio University, 5322 Endo, Fujisawa-shi, Kanagawa 252-0882 Japan, Email: makikon@sfc.keio.ac.jp.

1 Introduction

Month of birth affects one's well-being in not only school, but also in adulthood as it influences relative age in a grade cohort. Many papers have documented that relatively younger students in a given cohort perform more poorly in academic tests (Bedard and Dhuey (2006)), playing sports (Helsen et al. (2005)), and other school activities (Dhuey and Lipscomb (2008)). In addition, some studies have reported that month of birth affects long-term outcomes including university enrollment (Bedard and Dhuey (2006)), labor earnings (Kawaguchi (2011)), and top management (Du et al. (2012)) and political positions (Muller and Page (2016)).

While the negative consequences of being among the youngest in class have been well documented, what remains less known is how children and their parents respond to the negative birth month shocks beyond delaying school enrollment (known as redshirting). Do parents of younger students in a cohort make a compensatory investment? How do younger students use their time outside of school? Teachers and peer students can also influence the skill formation of younger students. Do teachers provide younger students with the required support? Do school peers help younger students or bully them instead? These questions are important for understanding children's skill formation and how skill gaps by birth month change over time.

The objectives of this paper are to estimate the effects of month of birth on cognitive and noncognitive skills as well as high school quality, and to examine factors that are likely to influence children's skill formation. More specifically, we estimate the month-of-birth gaps in time use, prep school attendance, participation in sports, and the quality of relationships with teachers and peer students.

Our analysis is based on longitudinal data of all fourth to ninth graders in public schools in an anonymous province¹ in Japan in 2015-2018. The data include over one million student-year observations. The timing of school enrollment in Japan is strictly determined by birth date, which differs from many countries, such as the USA, but is similar to institutions in England, Norway, and Iceland. Because red-

¹The actual administrative term used by the government is prefecture, but we use province as it is more intuitive for most readers.

shirting is virtually impossible in Japan, only the month of birth determines whether a student is one of the youngest in his/her grade cohort. We estimate the month-of-birth effects by assuming that the month of birth is exogenous. We validate this assumption by showing that family socioeconomic status (SES) and birth outcomes such as birth weight are largely uncorrelated with the month of birth.

Our estimation results indicate that the month of birth affects both cognitive and noncognitive skills, and hence, the quality of high school. The youngest students in grade four have test scores about 0.35 standard deviations lower than those of the oldest students; however, these gaps in test score decrease to 0.13-0.18 by grade nine. Findings for noncognitive skills such as conscientiousness are particularly interesting. Consistent with existing psychological studies on adolescence, noncognitive skills tend to decrease with grade in our data. However, the youngest students in a given grade have noncognitive skills about 0.1 standard deviations lower than the oldest students in the same grade. Because noncognitive skills do not improve with grade, the gap between the oldest and youngest students in the same grade suggests that relative age influences the formation of noncognitive skills. In addition, this month-of-birth gap in noncognitive skills remains nearly constant over time, which is in contrast with the dynamics of the cognitive skill gap.

We then examine how younger students and their parents respond to their disadvantages, and find that the youngest students in the ninth grade work 0.3 more hours per week outside of school than the oldest students. They also read for more hours and are more likely to attend a prep school. Overall, younger students and their parents make compensatory investment in cognitive skills; however, they spend 0.2 fewer hours per week on sports, arts, and music, which may lead to lower noncognitive skills, in light of the findings of psychological studies (Ishihara et al. (2018); Moreno et al. (2011)). Another important factor that may influence children's noncognitive skills is interpersonal relationships. Our estimates indicate that the youngest students suffer about 0.06 standard deviations poorer relationships with their teachers and 0.09-0.12 standard deviations poorer relationships with their school peers, which may lead to lower noncognitive skills for younger children (Leary and Baumeister (2000); Reitz et al. (2016); Wu et al. (2018)). Poor interpersonal relationships may lead to persistent month-of-birth effects on noncognitive

skills.

Our finding regarding compensatory skill investment has implications for efficiency and equity. Younger children and their parents seem to decide on skill investment through comparisons with exactly the same grade cohort; however, they will be compared with more diverse individuals when they enter the labor market. Compensatory skill investment may be individually optimal given their environment, but it may be suboptimal in the long run. The existing school cutoff date may distort students' and their parents' skill investment behavior. Our finding is also relevant for equity. Even though month-of-birth gaps may disappear in the long run, the cost of catching up with peers may be paid by younger students and their parents through their compensatory investment. Policies mitigating month-of-birth effects could correct the distortion in skill investment and improve equity.

Our results also provide a possible explanation for why month of birth has long-term consequences in some countries. For example, Kawaguchi (2011) finds that the youngest men in a grade cohort in Japan earn 4% less income when aged 30-34 years. Although month-of-birth effects on cognitive skills may disappear because of maturation and compensatory skill investment, less sports and music experience, as well as poorer interpersonal relationships with teachers and peers, may harm the long-term development of noncognitive skills, which might account for the lower earnings seen in adulthood. Although we do not observe noncognitive skills after grade nine in our data, this hypothesis is consistent with findings by Chetty et al. (2011), Heckman et al. (2013), and Baker et al. (2019), who find that interventions during childhood affect noncognitive skills in adulthood.

The rest of the paper is structured as follows. In Section 2, we outline the empirical strategy to identify month-of-birth effects and validate our identifying assumptions. In Section 3, we describe our data. In Section 4, we estimate month-of-birth effects on skills and high school quality. In Section 5, we estimate month-of-birth gaps in time use and interpersonal relationships. In Section 6, we discuss the results in the context of the literature. We conclude in Section 7. The appendix includes an extended literature review, details of the data, and additional analysis.

2 Empirical Strategy

2.1 Econometric Model

The school year in Japan runs from April to March. Children who have their sixth birthday on or before April 2 enter the first grade of elementary school of that year, which implies that the youngest children in a given grade were born in March and the oldest in April.² This law on school entry is strictly enforced, and hence, delaying school entry or redshirting is exceptionally rare. Because grade retention is also extremely uncommon, the timing of school entry and grade promotion in Japan is therefore completely determined by birth date. These rules are not quite common among OECD countries; however, England, Iceland, and Norway also have a school entry rule similar to that in Japan (Bedard and Dhuey, 2006).

We estimate gaps between the oldest and youngest students within the same grade in several outcomes using the following econometric model, which is estimated separately for each grade,

$$Y_{it} = \beta_0 + \beta_1 AGE_{it} + \beta_2 AGE_{it}^2 + \beta_3 X_{it} + \varepsilon_{it}, \quad (1)$$

where Y_{it} is an outcome for student i in year t , AGE_{it} is age at test measured in months, X_{it} is a vector of control variables, including gender, number of books at home as a proxy for SES, year, and school fixed effects, and ε_{it} is an error term. The variable AGE_{it} is normalized so that the youngest students in a given grade have the value of 0 while the oldest students have the value of 11. Using the estimated model, we then calculate the estimated gap between the oldest and youngest students, which is given by

$$\widehat{\Delta Y} = \hat{\beta}_1 \cdot 11 + \hat{\beta}_2 \cdot 11^2. \quad (2)$$

Crawford et al. (2014) argue that the gap is influenced by three factors. First, it

²Strictly speaking, students born on April 1 are the youngest and those born on April 2 are the oldest. Our data include the month of birth, not the exact date of birth, and hence, we refer to April-born students as the oldest. It is true that this group includes those born on April 1, who are actually the youngest, but we believe this classification error is negligible.

captures the absolute age or age-at-test effect. Because all of our outcomes are measured in April, students born in different months take tests at different absolute ages. Second, this parameter also captures the effect of age at school entry. Learning at school may not be effective if children are too young to study at school, and hence, age at school entry may have consequences for student outcomes at an older age. These two factors cannot be disentangled because age at test is the sum of the school starting age and the length of schooling (i.e., perfect collinearity). The third factor is relative age. In theory, we can identify a relative-age effect separately from the effects of absolute and school starting age. In practice, they are highly correlated, and hence, we do not attempt to identify the relative age effect separately.

The identifying assumption is conditional mean independence, i.e., $E(\varepsilon_{it} | Age_{it}, X_{it}) = E(\varepsilon_{it} | X_{it})$. One might be concerned that parents that are enthusiastic for education may time the birth to give their child the advantage of being among the oldest in his/her grade, which leads to endogeneity bias. We discuss this issue in the following subsection.

2.2 Assessing Exogeneity of Month of Birth

Our main data set is drawn from provincial skill assessments in an anonymous province near Tokyo, and covers nearly all students from grades four to nine in public schools for 2015-2018. The number of student-year observations is over one million. Although the data are rich in terms of coverage of students, the main limitation is that they include very few family/parental background variables, which prevents us from assessing exogeneity of month of birth directly. To overcome this limitation, we conduct two sets of exercises. In the first set, vital statistics that record month of birth, parental characteristics, and birth outcomes for all births in the province for the birth cohort in the main data set, i.e., all births from April 2000 to March 2011. In the second set, we use administrative data from a municipality in the province for 2015 and 2018. These data include information on whether a family is on welfare, receives an education subsidy as a low-income family, and/or is a single-parent household.

Figure 1 presents the birth month shares after adjustment for unequal number

of days in each month using vital statistics. Specifically, we divide the unadjusted relative frequency by the number of days and then multiply by 30. If all births are completely random, the share is 8.3% for each month. Thanks to a large number of observations, the share is statistically significantly different from 8.3% for some months, but the magnitudes of the differences are small. The biggest deviation from the average is found for September, with a share of 8.7%. Importantly, the gaps in the shares of March-born (youngest) children and April-born (oldest) children are tiny (7.9% and 8.0%, respectively).

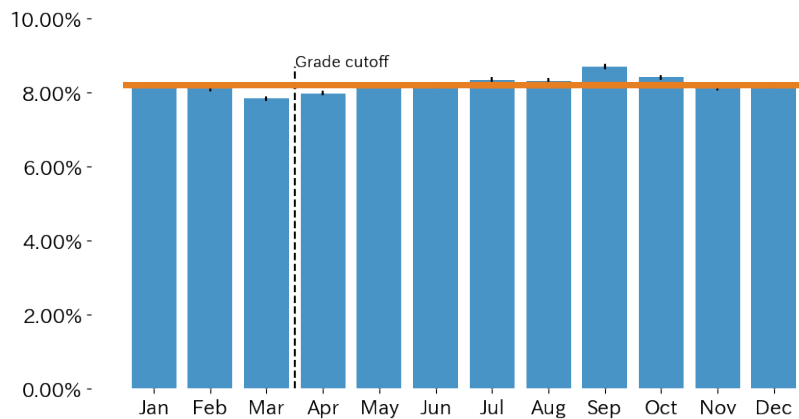


Figure 1: Relative Frequency of Birth Month (Adjusted for Number of Days)

Source: Vital Statistics from April 2000 to March 2011.

Note: The relative frequency of birth months is shown. We adjust for differences in the number of days across months as follows; we divide the unadjusted relative frequency by the number of days in the month and then multiply this by 30. The horizontal bar is drawn at 8.3%, which is the relative frequency of a birth month under the assumption that births are completely random. April-born children are the oldest, while March-born children are the youngest in a given grade.

We next examine the association between birth month and parental characteristics as well as birth outcomes. We regress parental characteristics and birth outcomes on age, age squared, and birth-year dummies using equation (1) and estimate the gaps between March- and April-born children using equation (2). Because we test multiple hypotheses simultaneously, we adjust the p-values throughout the paper using the method proposed by Benjamini and Hochberg (1995), which controls for a false discovery rate for independent hypotheses and provides more power than

the Bonferroni correction. Table 1 reports the results with adjusted p-values for 15 hypotheses. The first set of outcomes is parents' characteristics, such as age, marital status, and household occupation. The average age of mothers is 29.941 years and that of fathers is 32.053 years. Mothers and fathers of April-born (oldest) children are 0.107 and 0.125 years younger, respectively, than those of March-born (youngest) children. Although the vital statistics do not record parents' education, we expect that parents' age and education are positively correlated because Census and other statistics indicate that educated women tend to have their first child at an older age. One might expect that educated mothers try to time birth so that their child is born in April instead of March, but this is not the case. We find no significant gap by birth month in mother's marital status.

Occupation of household head, which is usually father if present, is classified into one of seven categories: farmer, self-employment, regular employment at a small firm, regular employment at a large firm, non-regular employment, none, and unknown. We say a firm is small if it hires fewer than 100 employees and large if it hires 100 employees or more. Large employers tend to pay higher wages and offer generous fringe benefits, although healthcare is universal in Japan. Compared with regular employment, non-regular employment is lower paid and unstable. We find no statistically significant gap in the fractions of farmers, self-employment, regular employment at a small firm, regular employment at a large firm, and unknown. More nonregular employment is observed among families of April-born (oldest) than among March-born (youngest) children, which is consistent with the fact that parents of April-born are younger. We also find that joblessness is slightly less prevalent among families of April-born than March-born children.

We also examine children's health at birth. Currie and Schwandt (2013) find that not only parental characteristics, but also seasonal environmental factors, such as flu prevalence, influence birth outcomes and may eventually lead to poor child development. Motivated by their findings, we examine the association of birth month with birth weight, gestation, and other birth outcomes. The regression results indicate no association between birth month and these infant health measures, being singleton, and gender; however, April-born (oldest) children are about one percentage point less likely to be a first child (mean, 49.2%). According to the literature,

Table 1: Gaps in Parental Characteristics and Birth Outcomes Between April- and March-born Children

Dependent variables	Mean	Estimates	Adjusted p-values
Parental characteristics			
Mother's age	29.941	-0.107 (0.021)	0.000
Father's age	32.053	-0.125 (0.023)	0.000
Unmarried (%)	1.910	-0.059 (0.072)	0.561
Occupation of household head			
Farmer (%)	0.690	-0.002 (0.039)	0.953
Self-employed (%)	7.185	-0.186 (0.106)	0.199
Regular employment (small firm) (%)	34.691	-0.033 (0.238)	0.953
Regular employment (large firm) (%)	41.608	-0.204 (0.179)	0.425
Non-regular employment (%)	8.749	0.499 (0.121)	0.000
No job (%)	1.723	-0.155 (0.050)	0.005
Unknown (%)	5.354	0.081 (0.192)	0.840
Birth outcomes			
Birth weight (gram)	3016.116	2.383 (1.909)	0.398
Gestational age (week)	38.897	-0.002 (0.009)	0.913
Singleton (%)	97.951	-0.076 (0.082)	0.528
First child (%)	49.229	-1.007 (0.259)	0.000
Girl (%)	48.710	-0.312 (0.239)	0.398
No. obs.	569983		

Source: Vital Statistics from April 2000 to March 2011.

Note: We regress parental characteristics and birth outcomes on age, age squared, and birth-year dummies using equation (1) and estimate the gaps between March- and April-born children using equation (2). Occupation of household head (father if present) is reported. We say a firm is small if it hires 1-99 employees and large if it hires 100 employees or more. Standard errors are in parentheses. Because we test 15 hypotheses simultaneously, we adjust the p-values for multiple testing using the method proposed by Benjamini and Hochberg (1995), which controls for a false discovery rate.

first-born children tend to perform better than their younger siblings. Although the magnitude is small, this first-born advantage is likely to bias downward, instead of upward, our estimated month-of-birth effects, if any.

Our second set of exercises use administrative data from a municipality in the anonymous province that conducted the skill assessment (our main data). These data include whether the family is on welfare, receives an education subsidy for low-income households, and is a single-parent household for 11,942 students in 2015 and 2018. We regress these family SES variables on age, age-squared, and grade dummies and then calculate the gap between April- and March-born students. The p-values are adjusted using the method proposed by Benjamini and Hochberg (1995) as there are three hypotheses. The results are reported in Table 2. We do not find any significant correlation between birth month and these family SES measures. To determine the extent to which we can generalize these results for the municipality to the whole province, we compare key variables between the two in Appendix B.1. Most importantly, little statistical and economic difference is seen between the two in terms of average family SES, which is proxied by the number of books in our main data set.

Table 2: Gaps in Family Background Between April- and March-born Students

	Mean	Estimates	Adjusted p-values
Education subsidy (%)	14.797	0.041 (1.214)	0.973
Single parent (%)	5.393	1.030 (0.727)	0.470
Welfare (%)	0.729	0.217 (0.265)	0.617
No. obs.		11942	

Source: Administrative data from a municipality in the anonymous province that conducted the skill assessment.

Note: We regress the family SES measures on age, age squared, and birth-year dummies using equation (1) and estimate the gaps between March- and April-born children using equation (2). Standard errors are in parentheses. We adjust the p-values for multiple testing using the method proposed by Benjamini and Hochberg (1995), which controls for a false discovery rate, as there are three hypotheses.

Overall, we do not find clear evidence that the oldest children in a grade cohort have more advantages from parental characteristics and birth outcomes than the youngest children. A few statistically significant differences are observed by birth month, but these are mostly because we have a large number of observations, and the magnitude does not seem economically large enough to change our substantive findings in this paper. We admit that the oldest children may be at a slight disadvantage, if any, which implies that our estimates for month-of-birth effects may be biased downward if this is the case.

A possible explanation for why month of birth may be uncorrelated with parental SES is that parents have a limited ability to time the birth. It is true that delaying birth for a few days by cesarean section is feasible (Shigeoka (2016)), but changing the birth date by more than a few days may not be straightforward. Parents might want to time the conception so that their child is born in April, but waiting for the right time to have an April-born baby can be costly. Given that the average child-bearing age is high, at 29.941 years, postponing conception can lower the chance of pregnancy.

3 Data

3.1 Overview

Our main data set is derived from standardized achievement tests and accompanied questionnaires that measure noncognitive skills and ask about student life in and outside the school. The tests have been administered by the school board in an anonymous province in a suburb of Tokyo every April since 2015, which is the beginning of the academic year. Virtually all students in grades four through nine in public schools in the province take this test and answer the questionnaire, although one municipality did not participate in the skill assessment. Only 2-3% of them did not take the exam for reasons such as illness.³ In total, 708 elementary and 356 junior high schools in 62 municipalities are included in the data set. Overall, 99.3%

³This participation rate is comparable to that for National Assessments of Academic Ability administered by the Ministry of Education, Culture, Sports, Science and Technology (MEXT).

of the elementary schools and 93.0% of the junior high schools are public in the province; only a few are private institutions.

Table 3 reports the descriptive statistics. The data cover the period 2015-2018 and include about 1.1 million student-year observations. For each grade, we have about 180,000 observations when pooled across years. From one year to the next, about 2% of students exit from the data set because they move out of the province or enter private schools. The attrition rates are 2.25%, 1.96%, 7.88%, and 2.81% for the transitions from grades four to five, five to six, six to seven, and seven to eight, respectively. The attrition rate is notably high when some students enter selective private junior high schools from grade seven. Because these students are likely to have higher cognitive and noncognitive skills, the results for primary school students (grades from four to six) and those for junior high school (grades from seven to nine) may not be fully comparable. However, we do not expect an attrition bias for other grades, and hence, the results are comparable across grades within primary and junior high school. In fact, our results do not change when we restrict our sample to individuals who participate in the skill assessment in all years from 2015 to 2018.

Although we have little information on family background, we use the number of books at home as a proxy for family SES because questions about parental income or educational background would have resulted in a low response rate and/or substantial measurement error. It is true that this is potentially an outcome influenced by educational achievement, but it is commonly used in sociology and education research.⁴ As reported in Table 3, 10.896% of students have 0-10 books at home in the period 2016-8, but this variable is unavailable for 2015.

⁴TIMSS (Woessmann, 2008) and PISA (Ammermueller, 2005) also use the number of books at home as a measure of family SES. Kawaguchi (2016) shows that the number of books at home is correlated with parental income and educational background in Japan.

Table 3: Descriptive Statistics

Count			
	All	1096062	
	Grade 4	179820	
	Grade 5	182261	
	Grade 6	186091	
	Grade 7	179281	
	Grade 8	182818	
	Grade 9	185791	
Demographics			
	Female (%)	49.658	
	0-10 books at home (%)	10.896	
		Mean	Std. Dev.
Cognitive skills			
	Math	0.192	1.344
	Japanese	0.233	1.666
	English	0.461	1.345
Noncognitive skills			
	Conscientiousness	3.734	0.628
	Self-control	2.649	0.438
	Self-efficacy	2.247	0.490
Time use & activities			
	Weekly hours of studying outside school	9.639	6.404
	Weekly hours of reading	4.153	3.786
	Prep school participation rate	0.584	0.493
	Weekly hours of playing outside and sports	3.989	3.028
	Weekly hours of arts, music, and sports	1.818	1.842
Interpersonal relationships			
	With teachers	2.430	0.600
	With peers	2.410	0.740

Source: Provincial Skill Assessment 2015-2018.

Note: Noncognitive skills, number of books at home, relationship with peers are available for 2016-2018. Relationship with teachers is available for 2017-2018. Weekly hours of reading, hours of playing outside and sports, and hours of arts, music, and sports are available only for 2015.

All other variables are available for 2015-2018.

3.2 Variable Definitions

3.2.1 Cognitive Skills

All students in our data set took math and Japanese tests, and the eighth and ninth graders also took an English test for 2015-2018. Employing item response theory (IRT; see Embretson and Reise (2000) for details), the tests are designed to enable comparisons of scores across different grades and test dates. IRT is implemented by including a few identical questions for anchoring in tests for different grades and years. This methodology is also used by the Programme for International Student Assessment (PISA) and TIMSS. In our data, test scores are normalized using the mean and standard deviation of the sixth graders for math and Japanese and those of the eighth graders for English in 2015. Consequently, the mean and standard deviation for the pooled sample reported in Table 3 are different from zero and one.

3.2.2 Noncognitive Skills

In a 40-minute questionnaire following the achievement test, the provincial school board has collected information on students' noncognitive skills, including conscientiousness, self-control, and self-efficacy, since 2016. To reduce the students' test burden, they measure one of the three noncognitive skills for each birth cohort. Figure 2 illustrates the survey structure. Consider the cohort who was in grade four in 2016 (born in 2006/2007). The school board measured their self-control from 2016 to 2018, but not other noncognitive skills. Similarly, the cohort who was in grade five in 2016 (born in 2005/2006) answered questions to measure their self-efficacy only. This survey structure is intended to measure the time evolution of noncognitive skills. As such, each of the noncognitive skills is measured by an identical set of questions across grades and survey years. The questions for noncognitive skills were translated into English, and are listed in Table 9 in the appendix.

Conscientiousness is defined as "the propensity to follow socially prescribed norms for impulse control, to be goal directed, to plan, and to be able to delay gratification" (Roberts et al., 2009) and is known as the most robust predictor of academic achievement among the Big Five personality traits.⁵ Students answered

⁵For example, Poropat (2009) meta-analysis on the relation among the Big Five personality traits

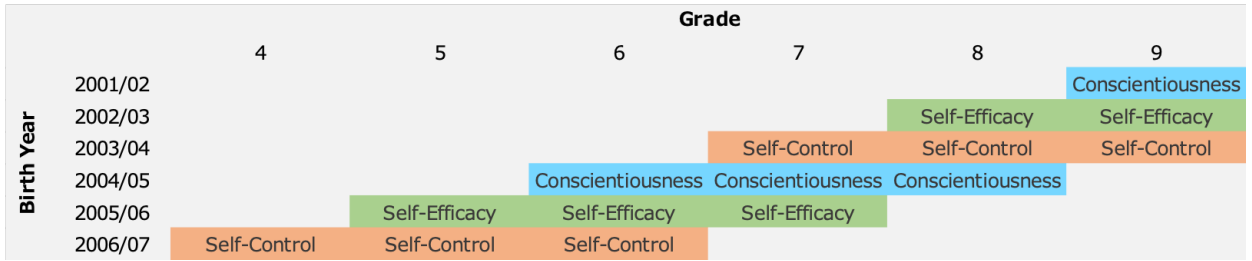


Figure 2: The Survey Structure for Noncognitive Skills

Note: The figure illustrates the structure of the survey for noncognitive skills. Only one noncognitive skill is measured for each birth cohort. Consider the cohort who was grade four in 2016 (born in 2006/2007). The school board measured their self-control from 2016 to 2018, but not other noncognitive skills.

13 questions about conscientiousness on a five-point Likert scale.

Self-control is a psychological state defined as “the tendency to regulate impulses and resist immediately rewarding temptations in the service of long-term goals” (Duckworth et al., 2013). Moffitt et al. (2011) show that childhood self-control predicts physical health, substance dependence, personal finances, and criminal offences in adulthood. Our measure of self-control is based on students’ answers to eight questions developed by Tsukayama et al. (2013). These questions are designed specifically for academic and school contexts.

Self-efficacy is defined as “people’s judgments of their capabilities to organize and execute courses of action required to attain designated types of performances (Bandura, 1986). Previous studies (Pintrich et al., 1991; Pajares and Miller, 1994; Pajares and Kranzler, 1995; Pajares and Graham, 1999) find that self-efficacy is strongly correlated with academic test scores, particularly with math.

For each of the three noncognitive skills, we sum all the relevant responses after correcting reverse-coded items. Table 3 reports the mean and standard deviation of

reveals that the both raw and partial correlations between conscientiousness and GPA are almost as large as those between GPA and IQ. Barbaranelli et al. (2003) examine the Big Five using factor analysis for adulthood to childhood to measure the personality ratings of children. They carefully investigate whether the children’s self-reporting reliably replicates parental and teacher descriptions of children’s personality and validate their scale as useful for measuring the Big Five personality traits through self-reporting by children aged 9-13 years.

the raw scores for each noncognitive skill for the pooled sample.

3.2.3 Human Capital Investment and Interpersonal Relationships

The questionnaire also asks about students' time use, prep school participation, and quality of relationships with teachers and peers. Time use and prep school participation are direct measures of human capital. The quality of interpersonal relationships is also relevant for human capital formation because teachers and peers can influence a student's behavior and skills.

We construct weekly hours of study outside of school for years 2015-2018 by combining average daily hours of study on weekdays and on weekends. The original response is provided as an interval, but we construct a point estimate by taking the midpoint of each interval and assuming 21×1.25 hours if students responded that they study 21 hours or more per week. Only in 2015, weekly hours are measured for reading, playing outside and sports, and organized out-of-school activities, including sports, music, and arts. Again, interval responses are converted into point estimates.

We also use whether a student attends a prep school as a measure of educational input for 2015-2018. Attending a prep school is common in Japan and other East Asian countries. Students attend a prep school to improve their academic skills and prepare for school entrance exams. While only about 5% of sixth graders take entrance exams for private junior high schools, nearly all ninth graders take entrance exams for high schools.⁶ The average annual expenditure on prep schools in 2016 was 218,000 Japanese yen ($\approx 2,180$ USD) for public elementary students, and 301,200 yen ($\approx 3,012$ USD) for public junior high school students (MEXT, 2017).

The means and standard deviations of these variables are reported in Table 3. Students study 9.639 hours per week outside of school and read 4.153 hours per week. Attendance at prep school is common, with 58.4% of students enrolled. Students play outside and/or sports for 3.989 hours per week and participate in organized arts, music, and sports activities outside of school for 1.818 hours per week.

⁶The high school enrollment rate was 98.8% in 2018.

To measure the quality of relationships with teachers, we use three questions asked in 2017-2018: “Did your teacher(s) assist you adequately?”, “Did your teacher(s) recognize your strengths?”, and “Did your teacher(s) help you understand content in class that you found confusing?”. To measure the quality of relationships with peer students, we use the question “Did your friend(s) recognize your strengths?” asked in 2016-2018. Students responded to these questions on a four-point scale. The means and standard deviations of these variables are reported in Table 3.

4 Month-of-Birth Effects on Skills and High School Quality

4.1 Cognitive Skills

We begin by examining graphically how average test scores change with age. Figure 3 shows the age-profiles for math, Japanese, and English test scores. We note two important features. First, average test scores increase with age at a decreasing rate. In a given grade, older students achieve higher test scores than do younger students, but the gap between younger and older students is smaller for upper-grade students. Second, mean test scores rise discontinuously at the grade cutoff for some grades. The jump between eighth and ninth grades is particularly large, presumably because ninth graders work much harder to get ready for entrance exams for high school. In the following, the focus is on the first feature or within-grade skill differences, while the second feature of the data is extensively analyzed in Appendix C.4.

We regress test scores on age and estimate the difference between the oldest and youngest students (i.e., the average of the oldest minus that of the youngest) for each grade. Note that the test scores in the regression analysis are normalized for each grade, which facilitates a comparison of estimates with those in other studies. Column (1) of Table 4 shows the estimated gaps in the mean math test scores between the oldest and youngest students in a given grade, using a quadratic function with no control variables except for age (see equations (1) and (2)). The gap is 0.350 standard deviations for grade four, but decreases as students move through higher grades. The gap is 0.125 standard deviations for grade nine, and is statistically sig-

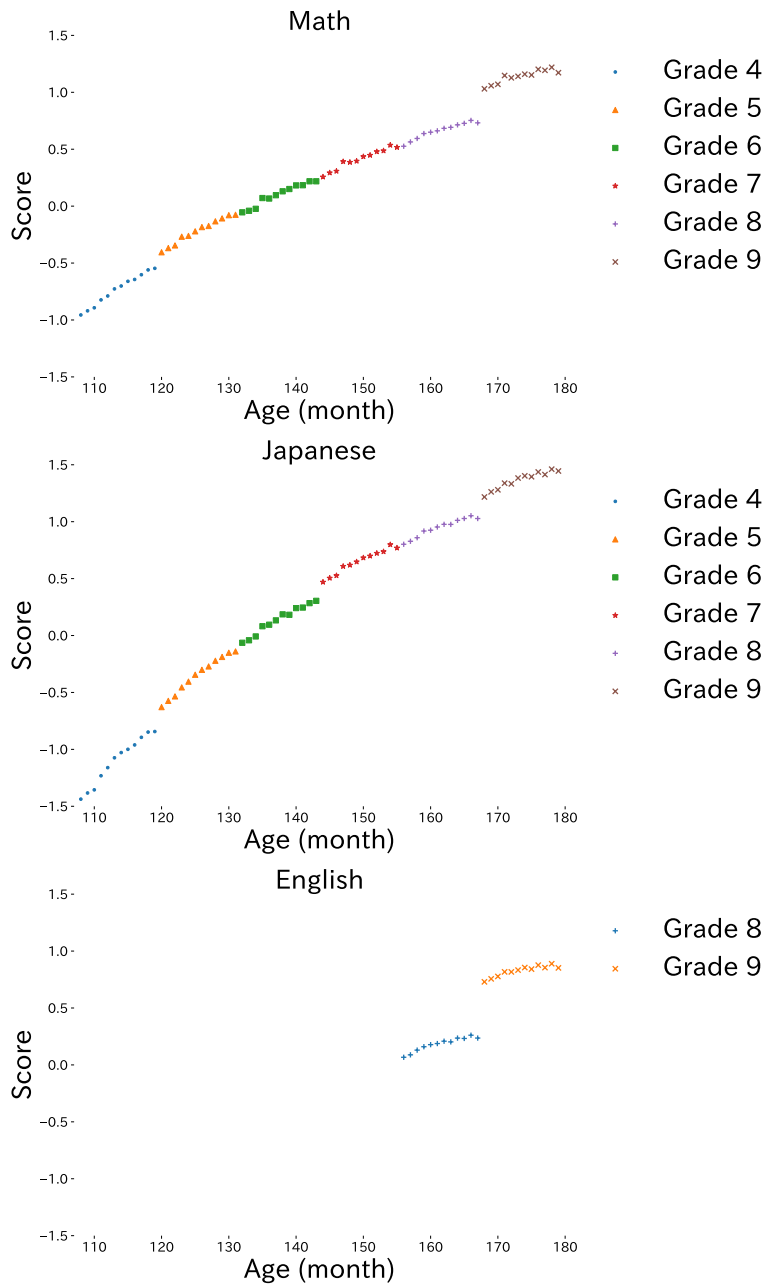


Figure 3: Mean Test Scores by Age

Source: Provincial Skill Assessment 2015-2018.

Note: Math and Japanese test scores are normalized so that the mean and standard deviation are 0 and 1 for sixth graders in 2015. English test scores are normalized similarly for eighth graders in 2015.

Table 4: Month-of-Birth Gaps in Test Scores

	Math (1)	Math (2)	Math (3)	Japanese (4)	Japanese (5)	Japanese (6)	English (7)	English (8)	English (9)
Grade 4	0.350 (0.008)	0.343 (0.009)	0.350 (0.011)	0.355 (0.008)	0.346 (0.008)	0.357 (0.011)			
Grade 5	0.297 (0.008)	0.290 (0.009)	0.293 (0.011)	0.319 (0.008)	0.320 (0.009)	0.321 (0.011)			
Grade 6	0.248 (0.008)	0.244 (0.009)	0.254 (0.011)	0.275 (0.008)	0.276 (0.008)	0.278 (0.011)			
Grade 7	0.228 (0.008)	0.225 (0.009)	0.224 (0.011)	0.245 (0.008)	0.242 (0.008)	0.248 (0.011)			
Grade 8	0.178 (0.007)	0.183 (0.008)	0.186 (0.011)	0.206 (0.007)	0.215 (0.008)	0.227 (0.011)	0.141 (0.007)	0.153 (0.008)	0.158 (0.011)
Grade 9	0.125 (0.008)	0.131 (0.008)	0.143 (0.011)	0.173 (0.008)	0.181 (0.008)	0.181 (0.010)	0.103 (0.008)	0.112 (0.008)	0.119 (0.010)
Control variables		✓	✓		✓	✓		✓	✓
Excl. Mar & Apr			✓			✓			✓

Note: The gaps between the oldest and youngest students are based on equations (1) and (2). Test scores in the regression analysis are normalized for each grade. The control variables include school fixed effects, survey year fixed effects, gender, and number of books at home. Standard errors are in parentheses. All estimates in this table are statistically significant at the 1% level when p-values are adjusted for multiple testing, using a method that controls for a false discovery rate (see Benjamini and Hochberg (1995)). We have 14 (6 + 6 + 2) hypotheses for each econometric specification.

nificant and smaller than that for grade four. Column (2) presents the estimated gaps when test year and school fixed effects, as well as students' gender and number of books at home, are controlled. The estimates are essentially the same as those in column (1). This is not surprising given that the month of birth is not correlated with most observed characteristics. Column (3) reports the estimates when March- and April-born students are excluded from the regression sample, which is motivated by the concern that parents may manipulate the timing of birth so that their children can take advantage of being the oldest in the class, rather than the youngest. The control variables are the same as those in column (2). The estimates are essentially the same as those in columns (1) and (2). Note that all estimates in Table 4 are statistically significant at the 1% level when p-values are adjusted by a method that controls for a false discovery rate (see Benjamini and Hochberg (1995)), as we test 14 (= 6 + 6 + 2) hypotheses simultaneously.

The estimated gaps in Japanese test scores are similar to those for math test scores. The gaps are 0.355 standard deviations for grade four, decreasing with grade, and 0.173 for grade nine. The estimates are robust across different specifications. The gaps in English test scores also decrease with grade, but are smaller than those in math and Japanese test scores. Given that English classes begin from grade seven, the smaller age gap in English test scores might imply that age effects are cumulative; by contrast, it could be because of the different nature of the subjects.

4.2 Noncognitive Skills

We plot the means of noncognitive skills by age in Figure 4. Note that students in different grades and survey years respond to identical questions that measure noncognitive skills, and hence, no gaps are driven by the questions themselves. The striking feature is that older students within the same grade have better noncognitive skills, despite the fact that noncognitive skills do not monotonically increase with grade. Indeed, noncognitive skills tend to decrease with grade, which contrasts with cognitive skills, which grow monotonically with grade.

Negative trends in psychological maturity during adolescence have been reported in previous psychological studies. For example, using a large-scale cross-sectional sample from English-speaking countries across the world, Soto et al. (2011) find that conscientiousness and other desirable psychological traits decrease when individuals transit from late childhood (age 10-12 years) into adolescence (age 13-17 years). Robins et al. (2002) find a similar negative trend for self-esteem. Denissen et al. (2013) provide a theoretical explanation in development psychology for these negative trends. On one hand, as children grow, their parents and teachers increasingly expect responsible behavior, which is referred to as a shift in the reference value. On the other hand, children change corresponding behavior in a desirable direction by internalizing new values and developing a behavioral repertoire to accommodate them. During late childhood and early adolescence, reference value shifts faster than children can change their behavior, which results in declining perceived maturity.

To focus on the gaps within grades, we normalize the noncognitive skills for

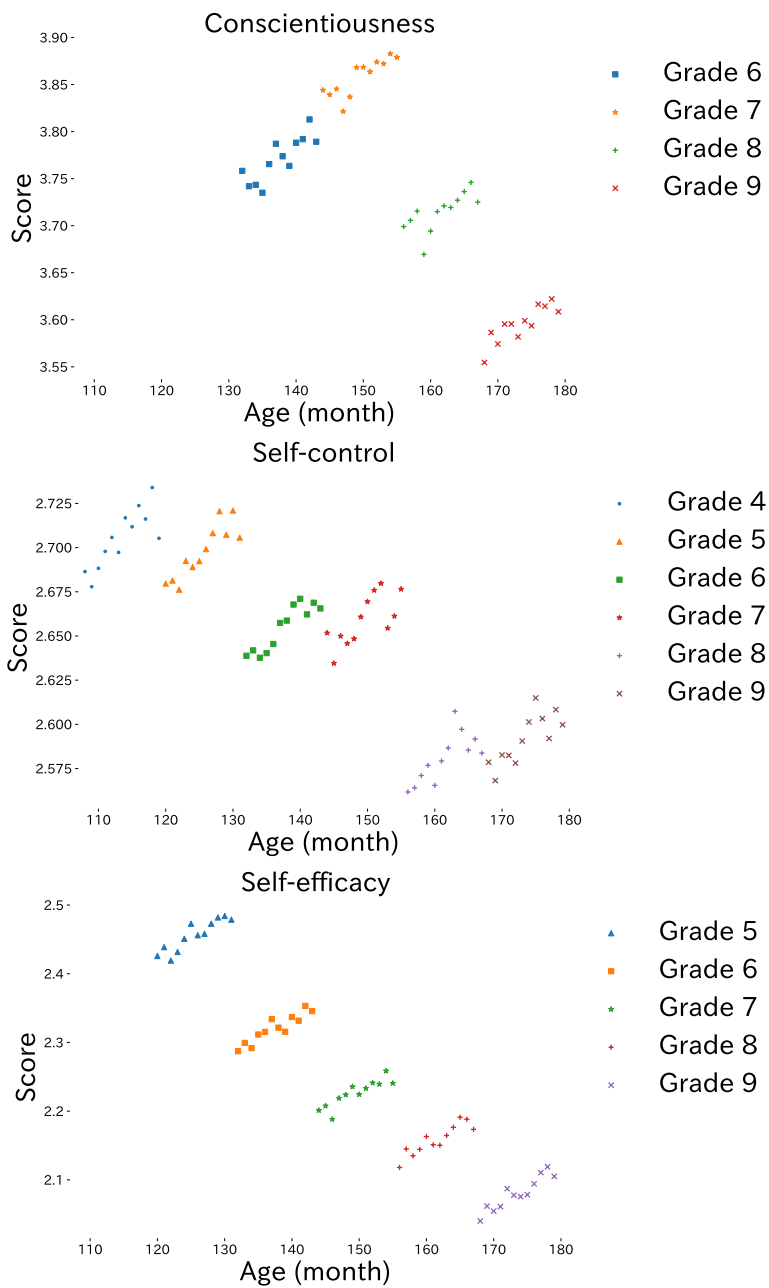


Figure 4: Noncognitive Skills by Age

Source: Provincial Skill Assessment 2016-2018.
 Note: Raw scores are reported.

Table 5: Month-of-Birth Gaps in Noncognitive Skills by Grade

	Conscientiousness		Self-control		Self-efficacy	
	(1)		(2)		(3)	
	Estimates	Adj. p-val.	Estimates	Adj. p-val.	Estimates	Adj. p-val.
Grade 4			0.092 (0.015)	0.000		
Grade 5			0.090 (0.015)	0.000	0.140 (0.016)	0.000
Grade 6	0.098 (0.015)	0.000	0.078 (0.015)	0.000	0.126 (0.015)	0.000
Grade 7	0.079 (0.015)	0.000	0.074 (0.014)	0.000	0.110 (0.015)	0.000
Grade 8	0.073 (0.015)	0.000	0.074 (0.014)	0.000	0.130 (0.015)	0.000
Grade 9	0.088 (0.014)	0.000	0.079 (0.014)	0.000	0.144 (0.015)	0.000
Control variables	✓		✓		✓	

Note: The gaps between the oldest and youngest students in a given grade are reported. In the regression analysis, noncognitive skill measures are normalized so that the mean is 0 and standard deviation is 1 for each grade. The control variables include school fixed effects, survey year fixed effects, gender, and number of books at home. Standard errors are in parentheses. We adjust p-values for multiple testing of 15 hypotheses, using a method that controls for a false discovery rate (see Benjamini and Hochberg (1995)).

each grade and then regress the normalized measures on age and other control variables. The results are presented in Table 5. Remember that we measure only one noncognitive skill for each birth cohort, and hence, not all noncognitive skills are measured for some grades. Column (1) shows the estimated gaps in conscientiousness between the oldest and youngest students. The estimate for grade six implies that the oldest students have 0.098 standard deviations better conscientiousness than the youngest students. The estimated gaps remain about the same in grades seven through nine. Column (2) presents the estimates for self-control. The estimated gaps are about 0.074-0.092 and are nearly constant across grades. Column (3) reports the estimates for self-efficacy. The estimates are 0.110-0.144 and do not change with grade. Note that the estimates are precisely estimated and the insignificant differences in the estimates across grades are not driven by large standard errors.

The month-of-birth effects show different patterns on noncognitive than on cog-

nitive skills. First, the effect size for noncognitive skills is about one-tenth of the standard deviations, which is smaller than the effect size for cognitive skills. Second, the month-of-birth effect on noncognitive skills remains nearly constant across grades, which is in contrast with the month-of-birth effect on cognitive skills, which decreases with grade.

4.3 Additional Analysis

In this subsection, we briefly present the main findings from a few additional sets of analysis. They are extensively discussed in Appendix C. In Section C.2, we estimate month-of-birth effects on high school quality using administrative data from the municipality in the province. We find that the oldest students enter better-quality high school than the youngest students in the same class. In Section C.3, we examine heterogeneity in the month-of-birth effects by gender and family SES, but find no evidence for heterogeneity in these respects. In addition, using quantile regressions, we estimate month-of-birth effects across different quantiles in an ability distribution. We find that the effects are larger in the right tail of the distribution, but the magnitude is modest.

5 Month-of-Birth Effects on Human Capital Investment and Interpersonal Relationships

We have shown that month of birth affects cognitive and noncognitive skills as well as high school quality significantly. In this section, we examine children's time use and activities outside of school to clarify whether and how the youngest students and their parents make compensatory investments. We also examine children's interpersonal relationships with their teachers and peers because these are likely to influence human capital formation both directly and indirectly.

5.1 Human Capital Investment

The top of Figure 5 shows the average weekly hours of study outside of school. While hours of study largely increase with grade, younger students in a given grade study longer hours, which suggests that relative age affects hours of study outside of school. A decline in hours of study is observed from grades six to seven, which might be explained by a composition change, as some students exit the survey from grade seven to enter private junior high school. The middle and bottom of Figure 5 show the average weekly hours of reading and enrollment rate for a prep school; both present patterns similar to hours of studying. Namely, they largely increase with grade, but younger students in a given grade are more likely to enroll in a prep school and read more.

The top panel of Figure 6 presents average hours per week playing sports and/or outside. Students in grades eight and nine play sports and/or outside more than elementary schoolers because school-based sports teams are more common in junior high school. According to the National Survey on Sports and Physical Activities 2018 by the Japan Sports Agency, the participation rate in school-based sports teams is 67.5% among junior high school students, compared with 24.1% among elementary school students. Although hours playing sports and/or outside change nonmonotonically with grade, younger students in a given grade do so for fewer hours than older students in any grade. The bottom panel of Figure 6 shows the average hours for organized art, music, and sports activities outside of school. Unfortunately, we are unable to disentangle hours for sports from those for art and music activities. The hours for these activities decrease with grade because students shift to prep school. Within each grade, younger students spend fewer hours on organized art, music, and sports outside of school.

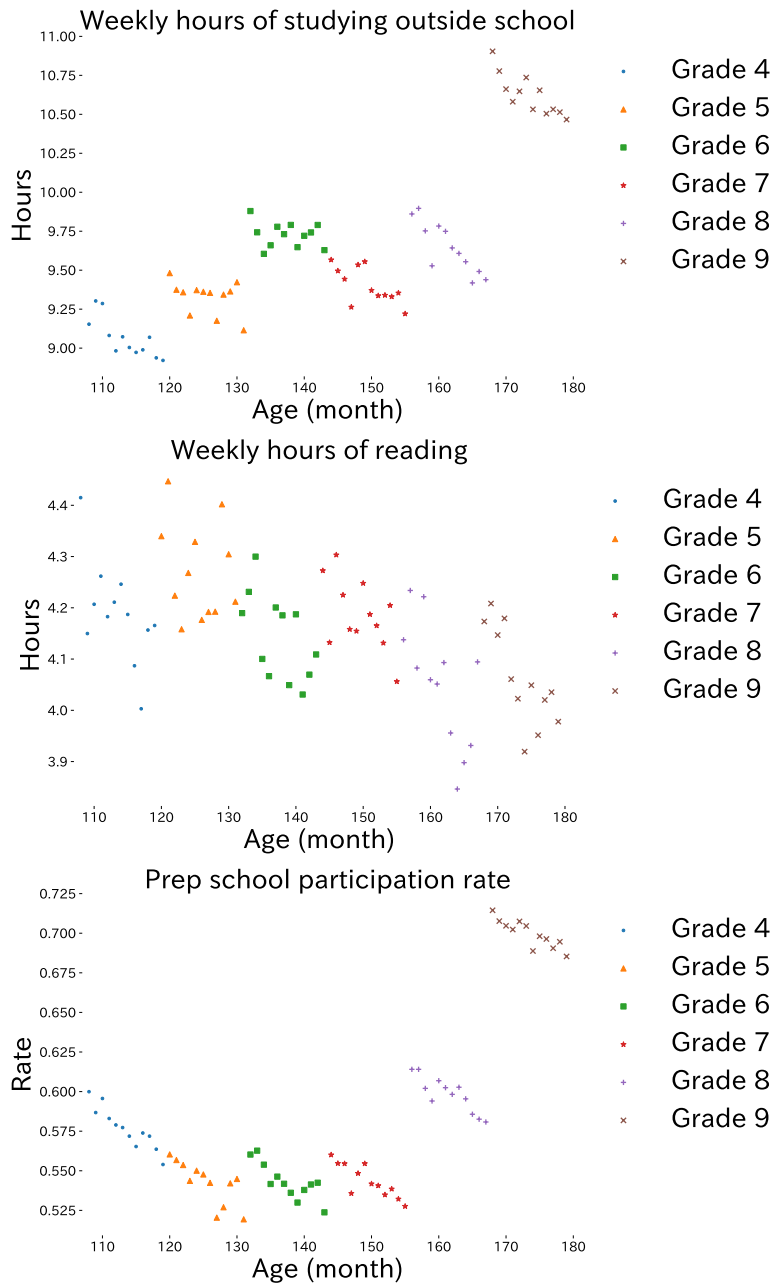


Figure 5: Human Capital Investment Outside of School (1)

Source: Provincial Skill Assessment 2015-2018.

Note: Hours of studying and prep school participation rates are available for 2015-2018. Hours of reading are available for 2015 only.

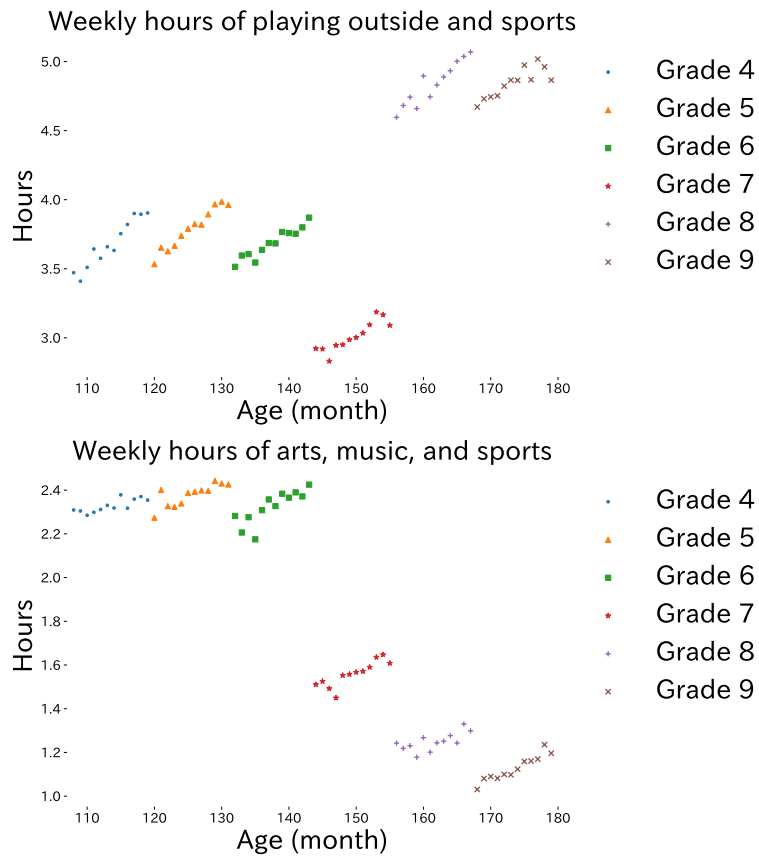


Figure 6: Human Capital Investment Outside of School (2)

Source: Provincial Skill Assessment 2015.

Note: Hours of playing sports and/or outside and hours of art, music, and sports are available for 2015 only.

Table 6: Month-of-Birth Effects on Human Capital Investment Outside of School

		Studying (1)	Reading (2)	Prep school (3)	Play outside/sports (4)	Arts, music, and sports (5)
Grade 4	Estimates	-0.361 (0.061)	-0.179 (0.069)	-0.039 (0.004)	0.517 (0.045)	0.077 (0.028)
	Adj. p-val.	0.000	0.012	0.000	0.000	0.007
Grade 5	Estimates	-0.145 (0.060)	-0.047 (0.065)	-0.032 (0.004)	0.459 (0.045)	0.140 (0.027)
	Adj. p-val.	0.018	0.470	0.000	0.000	0.000
Grade 6	Estimates	-0.048 (0.056)	-0.140 (0.065)	-0.028 (0.004)	0.307 (0.043)	0.192 (0.026)
	Adj. p-val.	0.402	0.034	0.000	0.000	0.000
Grade 7	Estimates	-0.241 (0.050)	-0.114 (0.061)	-0.027 (0.004)	0.296 (0.043)	0.153 (0.026)
	Adj. p-val.	0.000	0.066	0.000	0.000	0.000
Grade 8	Estimates	-0.402 (0.051)	-0.250 (0.072)	-0.028 (0.004)	0.441 (0.041)	0.085 (0.025)
	Adj. p-val.	0.000	0.001	0.000	0.000	0.001
Grade 9	Estimates	-0.295 (0.057)	-0.224 (0.074)	-0.022 (0.004)	0.285 (0.046)	0.164 (0.025)
	Adj. p-val.	0.000	0.003	0.000	0.000	0.000
Gender & school FE & year FE		✓	✓	✓	✓	✓
No. of books		✓		✓		

Source: Provincial Skill Assessment 2015-2018.

Note: Gaps between the oldest and youngest students are reported using equations (1) and (2). The variables Studying and Prep School are available for 2015-2018, while other dependent variables are available only for 2015. Standard errors are in parentheses. We adjust the p-values for multiple testing of 30 hypotheses, using a method that controls for a false discovery rate (see Benjamini and Hochberg (1995)).

We now turn to a regression analysis in which the variables for human capital investment are regressed on age, age squared, gender, and school and year fixed effects. The number of books at home is also controlled, if available. Table 6 reports the estimated gaps between the oldest and youngest students based on the regression model (see Equations 1 and 2) with p-values adjusted for multiple testing. Column (1) shows the results for hours of studying outside the home. The estimates are negative and statistically significant, except for grade six. Among ninth graders, the youngest students study 0.295 hours per week longer than the oldest students. We also find that the youngest students read at most 0.250 more hours and are 3.9 percentage points more likely to go to prep school than the oldest students in the same grade (see columns 2 and 3). By contrast, the youngest students spend at most 0.517 fewer hours per week playing sports and/or outside and 0.192 fewer hours per week on art, music, and sports activities outside of school.

5.2 Interpersonal Relationships

Figure 7 shows self-assessed relationships with teachers (top) and peers (bottom). Exactly the same questions are used across grades. The striking feature is that older students have better relationships with teachers and peers, even though the relationships with teachers worsen with grade and those with peers change nonmonotonically with grade. Again, the positive slopes within grade suggest that relative age influences relationships with teachers and peers.

Table 7 presents the estimated gap between the oldest and youngest students in a given grade, based on the regression model (see equations 1 and 2). In the regression, we normalize the outcome variables for each grade to facilitate interpretation. We find that younger students suffer 0.052-0.089 standard deviations poorer relationships with teachers and 0.096-0.127 standard deviations poorer relationships with peers. All of the estimates are statistically significant when the p-values are adjusted for multiple testing.

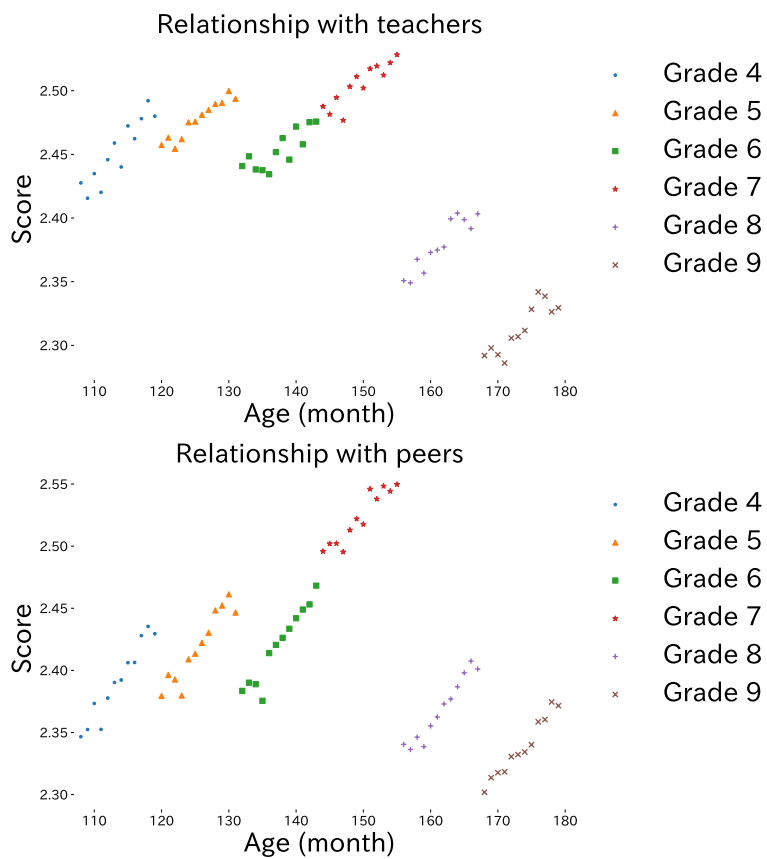


Figure 7: Quality of Relationships with Teachers and Peers

Source: Provincial Skill Assessment 2016-2018.

Note: Raw scores are reported. Relationship with teachers is available for 2017-2018, while that for peers is available for 2016-2018.

Table 7: Month-of-Birth Effects of Relationships with Teachers and Peers

		Teachers	Peers
		(1)	(2)
Grade 4	Estimates	0.089 (0.008)	0.122 (0.009)
	Adj. p-val.	0.000	0.000
Grade 5	Estimates	0.059 (0.008)	0.110 (0.008)
	Adj. p-val.	0.000	0.000
Grade 6	Estimates	0.052 (0.008)	0.127 (0.008)
	Adj. p-val.	0.000	0.000
Grade 7	Estimates	0.064 (0.009)	0.117 (0.009)
	Adj. p-val.	0.000	0.000
Grade 8	Estimates	0.078 (0.009)	0.100 (0.009)
	Adj. p-val.	0.000	0.000
Grade 9	Estimates	0.068 (0.008)	0.096 (0.009)
	Adj. p-val.	0.000	0.000
Control variables		✓	✓

Source: Provincial Skill Assessment 2016-2018.

Note: The gaps in the quality of interpersonal relationships between April- and March-born students are reported. The dependent variables are normalized for each grade to have a zero mean and a standard deviation of one. Control variables include school fixed effects, survey year fixed effects, gender, and number of books at home. Standard errors are in parentheses. The p-values are adjusted for multiple testing of 12 hypotheses, using a method that controls for a false discovery rate (see Benjamini and Hochberg (1995)).

6 Discussion

Our estimates indicate that the youngest students in grade four achieve 0.35 standard deviations lower test scores than the oldest students; however, by grade nine, this gap diminishes to 0.13-0.18 standard deviations. The magnitude of month-of-birth effects is similar to the previous estimates across different countries. The most comparable study to the current paper is Bedard and Dhuey (2006), who use TIMSS, which includes data from Japan. They find that the impact of month-of-birth is 0.25-0.31 standard deviations for grade four and 0.23 standard deviations for grade eight among Japanese students.

We also find that month-of-birth effects on noncognitive skills are significant, and the gap between the oldest and youngest students in the same grade is about 0.1 standard deviations. Interestingly, month-of-birth effects on noncognitive skills remain constant from grade four to nine. Only a few previous papers have estimated month-of-birth effects on noncognitive skills. These papers suggest that month-of-birth influences many, if not all, types of noncognitive skills.⁷ Among them, only Datar and Gottfried (2015) and Peña and Duckworth (2018) estimate month-of-birth effects on noncognitive skills for different grades. Datar and Gottfried (2015) find that the effect is significant for elementary school students, but not for those in grade eight in the USA. By contrast, Peña and Duckworth (2018) find that the effects on grit do not vary between ninth and twelfth graders (0.122 vs. 0.145 standard deviations, respectively) in Mexico. We add to this literature using data that cover a broader range of noncognitive skills for a longer period of time than in previous studies.

We depart previous contributions by examining human capital investment and interpersonal relationships, both of which are relevant for children's skill formation. We find evidence of compensatory investment by the youngest children and their parents. Namely, the youngest students study and read more hours and are more likely to go to a prep school than are the oldest students in the same grade. Although our data do not allow us to test whether these inputs improve test scores,

⁷See Crawford et al. (2014), Datar and Gottfried (2015), Lubotsky and Kaestner (2016), Page et al. (2017), and Peña and Duckworth (2018)

we find hours of study outside of school and prep school attendance to be positively correlated with test scores after controlling for student fixed effects (see Appendix C.5). Moreover, Stinebrickner and Stinebrickner (2008) find causal evidence that hours of studying improve the GPA of first year students at a US college. In a randomized controlled trial (RCT), Akabayashi et al. (2018) find that after-school education programs improve Japanese students' test scores.

Regarding noncognitive skills, we suspect that the youngest students' low participation at sports, music, and art activities and poor-quality relationships with peers and teachers exacerbate their disadvantage in noncognitive skills development relative to the oldest students. Psychologists argue that sports, music, and art improve noncognitive skills. For example, Ishihara et al. (2018) find a strong association between sports experience and noncognitive skills. Moreno et al. (2011) find causal evidence in an RCT that music training enhances noncognitive skills. In addition, psychologists find evidence that poor interpersonal relationships negatively affect noncognitive skills.⁸ In a recent study, Wu et al. (2018) find strong association between teacher-child relation and children's social skills. Reitz et al. (2016) find that popularity among peers predicts self-esteem, but the opposite is not true, which provides evidence in support of sociometer theory,⁹ which argues that self-esteem reflects an individual's perceptions of how they are viewed by others, mediated by self-perceived popularity. The current paper also finds that personal relationships are positively correlated with noncognitive skills after controlling for student fixed effects (see Appendix C.5), although this does not necessarily imply causal effects. The observed patterns of time use and interpersonal relationships across children with different birth months may explain why month-of-birth gaps in noncognitive skills do not disappear as students grow.

Our findings on time use and prep school attendance suggest that skill investment is distorted by the existing school cutoff. They imply that parents decide

⁸Davis (2003) synthesizes several psychological studies and finds that both attachment and motivation theories predict that teacher-child relationships improve children's social and cognitive outcomes.

⁹Leary and Baumeister (2000) develop the sociometer theory, which argues that self-esteem reflects an individual's perceptions of how they are viewed by others, mediated by self-perceived popularity.

on skill investment by comparing their children with exactly the same grade cohort; however, their children will be compared with more diverse individuals once they are out of school. Compensatory skill investment is an individually optimal strategy in the short run, but may be suboptimal from a long-term and/or societal perspective. Note that the same argument applies to older students because they may underinvest in cognitive skills. Hence, correcting the distortion may improve the aggregate human capital stock and benefit all students in the long run. While this issue has not been raised in the relative age literature, to the best of our knowledge, Kinsler and Pavan (forthcoming) also show evidence that parental investment is distorted by local information.¹⁰

Our results indicate that month of birth has negative consequences in the short run. First, the youngest students suffer poor relationships with teachers and peers, which is a serious concern. Second, the evidence for compensatory investment implies that younger students and their parents pay the cost to fill the month-of-birth gaps, although maturation may also be an important factor for convergence. Our results suggest that, even if part of the month-of-birth gaps disappear by adulthood, month of birth hurts the welfare of younger children in a given grade cohort. Indeed, Matsubayashi and Ueda (2015) find that month of birth affects the suicide rate among young Japanese aged 15-25 years.

Furthermore, these findings may account for the month-of-birth effects on adulthood outcomes reported in some countries. Specifically, noncognitive skill gaps may be one of key driving forces behind the month-of-birth effects on earnings at age 30-34 years reported by Kawaguchi (2011) using data from Japanese Labor Force Survey. Noncognitive skills not only help one acquire cognitive skills, but also have independent effects on labor market outcomes (Heckman et al., 2006). If month-of-birth effects on noncognitive skills remain in adulthood, they are likely to translate into an earnings gap. Of course, this is only a hypothesis, because our data do not follow individuals after grade nine¹¹; however it is consistent with Dhuey and

¹⁰They find that parents form a belief about their child's skill by comparing the child with those in the same school, rather than those of the same age.

¹¹The Preference Parameters Study of Osaka University measures noncognitive skills for Japanese adults, but the sample size is only about 2000, which is too small to detect the small gap in noncognitive skills found in this paper.

Lipscomb (2008), who find that relatively older students take leadership activities in high school. In addition, Chetty et al. (2011), Heckman et al. (2013), and Baker et al. (2019) find that noncognitive skills formed in childhood influence outcomes in adulthood. It is true that month-of-birth effects on cognitive skills may also be preserved into adulthood through a high school entrance exam (see Appendix C.2) and eventually influence earnings in adulthood, but our findings offer an alternative explanation for the persistent effect of month-of-birth.

7 Conclusion

We estimated month-of-birth effects on cognitive skills, noncognitive skills, and high school quality using a large longitudinal data set for students in grades four through nine. We find statistically and economically significant effects of month-of-birth on all of these outcomes. While the effects on cognitive skills diminish over time, those on noncognitive skills remain constant throughout elementary and junior high school.

We explored children's time use and interpersonal relationships, both of which are relevant for their skill formation. On one hand, we find evidence of compensatory investment in cognitive skills by younger students in a given cohort. Younger students spend more hours studying and reading and are more likely to attend private prep schools. On the other hand, they are less likely to participate in sports, music, and arts, and have poorer relationships with teachers and peers, all of which may harm the long-term development of noncognitive skills. This finding offers an alternative explanation for why month of birth affects adulthood outcomes in some countries, including Japan.

Our findings raise concern about efficiency and equity. Different time use and prep school attendance by birth month suggest that skill investment may be distorted for not only younger, but also older students from a long-term perspective. We also find that younger students suffer the disadvantage of poor-quality interpersonal relationships, which is caused by the arbitrarily determined school cutoff date. Policies that mitigate month-of-birth effects could increase the aggregate human capital stock and improve individuals' welfare. For example, age-adjusted

evaluation may be a feasible option, and is particularly relevant for high school and university entrance exams.

Our findings also have implications for the current policy debate regarding whether to change the school starting date in Japan from April to September. Due to the COVID-19 pandemic, the Japanese government closed schools from March to May 2020. Some politicians argue for deferring the school starting date from April to September commencing in 2021 to make up for the lost school days. One of the possible consequences is an increase in cohort size, because children born between April 2014 and August 2015 would start school together. This would increase the maximum age gap from 11 to 17 months and exacerbate the disadvantages of younger students in the cohort. If they change the school starting date, the government should minimize the growth of cohort size and provide necessary assistance to younger students.

There are at least two limitations of this paper. First, our data do not include skills and other outcomes in adulthood. Second, our data do not allow us to estimate the causal effects of time use and interpersonal relationships on skills because we are unable to find exogenous variations in these factors, although other papers have established causality for some. These issues should be addressed in future research.

Bibliography

- Akabayashi, Hideo, Hiroko Araki, and Ryuichi Tanaka. 2018. Effects of After-School Education Vouchers on Children's Academic and Behavioral Outcomes: Evidence from a Randomized Experiment. Tech. Rep. 48, University of Tokyo.
- Attar, Itay and Danny Cohen-Zada. 2018. The effect of school entrance age on educational outcomes: Evidence using multiple cutoff dates and exact date of birth. Journal of Economic Behavior & Organization 153:38–57.
- Baker, Michael, Jonathan Gruber, and Kevin Milligan. 2019. The long-run impacts of a universal child care program. American Economic Journal: Economic Policy 11 (3):1–26.

- Bandura, Albert. 1986. The explanatory and predictive scope of self-efficacy theory. Journal of social and clinical psychology 4 (3):359–373.
- Barbaranelli, Claudio, Gian Vittorio Caprara, Annarita Rabasca, and Concetta Pastorelli. 2003. A questionnaire for measuring the big five in late childhood. Personality and Individual Differences 34 (4):645–664.
- Bedard, Kelly and Elizabeth Dhuey. 2006. The persistence of early childhood maturity: International evidence of long-run age effects. The Quarterly Journal of Economics 121 (4):1437–1472.
- Benjamini, Yoav and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological) 57 (1):289–300.
- Black, Sandra E, Paul J Devereux, and Kjell G Salvanes. 2011. Too young to leave the nest? the effects of school starting age. The Review of Economics and Statistics 93 (2):455–467.
- Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. How does your kindergarten classroom affect your earnings? Evidence from project star. Quarterly Journal of Economics 126 (4):1593–1660.
- Crawford, Claire, Lorraine Dearden, and Ellen Greaves. 2014. 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) 177 (4):829–860.
- Currie, Janet and Hannes Schwandt. 2013. Within-mother analysis of seasonal patterns in health at birth. Proceedings of the National Academy of Sciences 110 (30):12265–12270.
- Datar, Ashlesha and Michael A Gottfried. 2015. School entry age and children, Äôs social-behavioral skills: Evidence from a national longitudinal study of us kindergartners. Educational evaluation and policy analysis 37 (3):333–353.
- Davis, Heather A. 2003. Conceptualizing the role and influence of student-teacher relationships on children’s social and cognitive development. Educational psychologist 38 (4):207–234.

- Dawid, Herbert and Gerd Muehlheusser. 2015. Repeated selection with heterogeneous individuals and relative age effects. Journal of Economic Behavior & Organization 116:387–406.
- Denissen, Jaap J.A., Marcel A.G. van Aken, Lars Penke, and Dustin Wood. 2013. Self-Regulation Underlies Temperament and Personality: An Integrative Developmental Framework. Child Development Perspectives 7 (4):255–260.
- Dhuey, Elizabeth and Stephen Lipscomb. 2008. What makes a leader? relative age and high school leadership. Economics of Education Review 27 (2):173–183.
- Dobkin, Carlos and Fernando Ferreira. 2010. Do school entry laws affect educational attainment and labor market outcomes? Economics of education review 29 (1):40–54.
- Du, Qianqian, Huasheng Gao, and Maurice D Levi. 2012. The relative-age effect and career success: Evidence from corporate ceos. Economics Letters 117 (3):660–662.
- Duckworth, Angela L, Eli Tsukayama, and Teri A Kirby. 2013. Is it really self-control? examining the predictive power of the delay of gratification task. Personality and Social Psychology Bulletin 39 (7):843–855.
- Dustmann, Christian, Patrick A Puhani, and Uta Schönberg. 2017. The long-term effects of early track choice. The Economic Journal 127 (603):1348–1380.
- Embretson, Susan E and Steven Paul Reise. 2000. Item response theory for psychologists. maheah.
- Fredriksson, Peter and Björn Öckert. 2014. Life-cycle effects of age at school start. The Economic Journal 124 (579):977–1004.
- Grenet, Julien. 2009. Academic performance, educational trajectories and the persistence of date of birth effects. evidence from france. Unpublished manuscript .
- Heckman, James, Rodrigo Pinto, and Peter Savelyev. 2013. Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. American Economic Review 103 (6):2052–86.
- Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. National Bureau of Economic Research 24 (3):1689–1699.

- Helsen, Werner F., Jan Van Winckel, and A. Mark Williams. 2005. The relative age effect in youth soccer across Europe. Journal of Sports Sciences 23 (6):629–636.
- Ishihara, Toru, Shigemi Sugawara, Yusuke Matsuda, and Masao Mizuno. 2018. Relationship between sports experience and executive function in 6-12-year-old children: independence from physical fitness and moderation by gender. Developmental Science 21 (3):1–13.
- Kawaguchi, Daiji. 2011. Actual age at school entry, educational outcomes, and earnings. Journal of the Japanese and International Economies 25 (2):64–80.
- . 2016. Fewer school days, more inequality. Journal of the Japanese and International Economies 39:35–52.
- Kinsler, Josh and Ronni Pavan. forthcoming. Local distortions in parental beliefs over child skill. Journal of Political Economy .
- Larsen, Erling Røed and Ingeborg F Solli. 2017. Born to run behind? persisting birth month effects on earnings. Labour Economics 46:200–210.
- Leary, Mark R and Roy F Baumeister. 2000. The nature and function of self-esteem: Sociometer theory. In Advances in experimental social psychology, vol. 32. Elsevier, 1–62.
- Lubotsky, Darren and Robert Kaestner. 2016. Dostills 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 53:194–206.
- Matsubayashi, Tetsuya and Michiko Ueda. 2015. Relative Age in School and Suicide among Young Individuals in Japan: A Regression Discontinuity Approach. PLOS ONE 10 (8):e0135349.
- Matta, Rafael, Rafael P Ribas, Breno Sampaio, and Gustavo R Sampaio. 2016. The effect of age at school entry on college admission and earnings: a regression-discontinuity approach. IZA Journal of Labor Economics 5 (1):9.
- McEwan, Patrick J and Joseph S Shapiro. 2008. The benefits of delayed primary school enrollment discontinuity estimates using exact birth dates. Journal of human Resources 43 (1):1–29.

- Moffitt, Terrie E, Louise Arseneault, Daniel Belsky, Nigel Dickson, Robert J Hancox, HonaLee Harrington, Renate Houts, Richie Poulton, Brent W Roberts, Stephen Ross et al. 2011. A gradient of childhood self-control predicts health, wealth, and public safety. Proceedings of the National Academy of Sciences 108 (7):2693–2698.
- Moreno, Sylvain, Ellen Bialystok, Raluca Barac, E. Glenn Schellenberg, Nicholas J. Cepeda, and Tom Chau. 2011. Short-term music training enhances verbal intelligence and executive function. Psychological Science 22 (11):1425–1433.
- Mühlenweg, Andrea M and Patrick A Puhani. 2010. The evolution of the school-entry age effect in a school tracking system. Journal of human Resources 45 (2):407–438.
- Muller, Daniel and Lionel Page. 2016. Born leaders: political selection and the relative age effect in the us congress: D. muller and l. page. Journal of the Royal Statistical Society: Series A (Statistics in Society) 179 (3):809–829.
- Nam, Kigon. 2014. Until when does the effect of age on academic achievement persist? evidence from korean data. Economics of Education Review 40:106–122.
- Page, Lionel, Dipanwita Sarkar, and Juliana Silva-Goncalves. 2017. The older the bolder: Does relative age among peers influence children's preference for competition? Journal of Economic Psychology 63:43–81.
- Pajares, Frank and Laura Graham. 1999. Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. Contemporary educational psychology 24 (2):124–139.
- Pajares, Frank and John Kranzler. 1995. Self-efficacy beliefs and general mental ability in mathematical problem-solving. Contemporary educational psychology 20 (4):426–443.
- Pajares, Frank and M David Miller. 1994. Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. Journal of educational psychology 86 (2):193.
- Peña, Pablo A. 2017. Creating winners and losers: Date of birth, relative age in school, and outcomes in childhood and adulthood. Economics of Education Review 56:152–176.

- Peña, Pablo A and Angela L Duckworth. 2018. The effects of relative and absolute age in the measurement of grit from 9th to 12th grade. Economics of Education Review 66:183–190.
- Pintrich, Paul R et al. 1991. A manual for the use of the motivated strategies for learning questionnaire (mslq). ERIC.
- Poropat, Arthur E. 2009. A meta-analysis of the five-factor model of personality and academic performance. Psychological bulletin 135 (2):322.
- Puhani, Patrick A. and Andrea M. Weber. 2007. Does the early bird catch the worm? Empirical Economics 32 (2-3):359–386.
- Reitz, Anne K., Frosso Motti-Stefanidi, and Jens B. Asendorpf. 2016. Me, us, and them: Testing sociometer theory in a socially diverse real-life context. Journal of Personality and Social Psychology 110 (6):908–920.
- Roberts, B. W., J. J. Jackson, J. V. Fayard, G. Edmonds, and J. Meints. 2009. Conscientiousness. In Handbook of individual differences in social behavior, ed. M. R. Leary & R. H. Hoyle. The Guilford Press, 369–381.
- Robins, Richard W., Kali H. Trzesniewski, Jessica L. Tracy, Samuel D. Gosling, and Jeff Potter. 2002. Global self-esteem across the life span. Psychology and Aging 17 (3):423–434.
- Shigeoka, Hitoshi. 2016. School entry rule and the timing of births. Tech. rep., Simon Fraser University.
- Soto, Christopher J., Oliver P. John, Samuel D. Gosling, and Jeff Potter. 2011. Age Differences in Personality Traits From 10 to 65: Big Five Domains and Facets in a Large Cross-Sectional Sample. Journal of Personality and Social Psychology 100 (2):330–348.
- Stinebrickner, Ralph and Todd R. Stinebrickner. 2008. The causal effect of studying on academic performance. B.E. Journal of Economic Analysis and Policy 8 (1).
- Tsukayama, Eli, Angela Lee Duckworth, and Betty Kim. 2013. Domain-specific impulsivity in school-age children. Developmental Science 16 (6):879–893.

- Tukiainen, Janne, Tuomas Takalo, and Topi Hülkkonen. 2017. Gender specific relative age effects in politics and football. Working Papers 94, VATT Institute for Economic Research.
- Wu, Zhongling, Bi Ying Hu, Xitao Fan, Xiao Zhang, and Juan Zhang. 2018. The associations between social skills and teacher-child relationships: A longitudinal study among Chinese preschool children. Children and Youth Services Review 88 (March):582–590.
- Zweimüller, Martina. 2013. The effects of school entry laws on educational attainment and starting wages in an early tracking system. Annals of Economics and Statistics/ANNALES D'ÉCONOMIE ET DE STATISTIQUE :141–169.

For Online Publication

A Extended Literature Review

This paper contributes to a large literature on the effects of month of birth. It has been widely documented that the younger children in a grade cohort perform more poorly in academic tests; examples include, but are not limited to, Bedard and Dhuey (2006), Puhani and Weber (2007), McEwan and Shapiro (2008), Mühlengeweg and Puhani (2010), Crawford et al. (2014), Nam (2014), Matta et al. (2016), Peña (2017), and Attar and Cohen-Zada (2018). For example, Bedard and Dhuey (2006) find that the youngest students of each cohort score 4-12 percentiles lower than the oldest students in grade four and 2-9 percentiles lower in grade eight, using Trends in International Mathematics and Science Study (TIMSS), which covers 19 Organisation for Economic Co-operation and Development (OECD) countries. The month-of-birth gaps in test scores seem to be largely explained by absolute age effects because the gaps diminish with grade. In fact, Black et al. (2011) find no month-of-birth effects on IQ tests measured around the 18th birthday. Nevertheless, little remains known about why month-of-birth effects decrease with absolute age. Maturation is obviously an important reason, but the youngest students' compensatory behavior can also explain the catching-up process, which has been paid scant attention in the literature.

Evidence for month-of-birth effects on noncognitive skills is mixed and more scarce than that for cognitive skills. Crawford et al. (2014) find that the youngest students value their own scholastic competence less than other students, but observe no month-of-birth effects on self-esteem and locus of control among UK students at age 8 years. Datar and Gottfried (2015) and Lubotsky and Kaestner (2016) find that the youngest students have weaker social-behavioral outcomes during kindergarten and elementary school in the USA, but the negative effect on young students disappears before high school. Page et al. (2017) examine Australian students and find that the youngest male students dislike competition more than do the oldest students, but observe no month-of-birth effects on risk attitude or self-confidence. Peña and Duckworth (2018) show that the youngest students in grades six and nine

in Mexico have less grit. Overall, month-of-birth effects are significant at least for some dimensions of noncognitive skills, and the effect tends to be stronger for lower-grade students. We contribute to the literature as our large panel data set with one million observations covers a broader range of noncognitive skills for a longer period than previous studies.

Several papers estimate month-of-birth effects on educational attainment and labor market outcomes, but the evidence is mixed. Black et al. (2011) and Larsen and Solli (2017) find no month-of-birth effects on prime-age earnings in Norway. Dobkin and Ferreira (2010) report similar results for California and Texas in the USA. Nam (2014) and Dustmann et al. (2017) find no month-of-birth effects on educational attainment or earnings in Korea and Germany, respectively.

Many other papers report significant month-of-birth effects on educational attainment and labor market outcomes. Most importantly for this paper, Kawaguchi (2011) finds that the youngest individuals in grade cohorts received 0.13 fewer years of education and earn about 4% less income at ages 30-34 years in Japan. Bedard and Dhuey (2006), Grenet (2009), Zweimüller (2013), Matta et al. (2016), and Peña (2017) find significant month-of-birth effects on university attainment and/or earnings in Canada and the USA, France, Austria, Brazil, and Mexico, respectively. It is also worth mentioning that Fredriksson and Öckert (2014) report significant month-of-birth effects on educational attainment and earnings of women and individuals with poorly-educated parents in Sweden. Lastly, month of birth also affects the chance of occupying highly competitive positions, including the CEOs of US S&P 500 companies (Du et al., 2012) and top political positions in the USA (Muller and Page, 2016) and Finland (Tukiainen et al., 2017).

While numerous papers document month-of-birth gaps in short- and long-run outcomes, little remains known about factors that mitigate or exacerbate these gaps. One of the most common explanations for why month-of-birth effects are persistent is early school tracking. The youngest children in a grade cohort are less likely to be selected for a higher quality academic program because they perform worse than their older peers, which exacerbates the gap between the oldest and youngest students. Dawid and Muehlheusser (2015) theoretically show that early school tracking can result in an erroneous selection. Fredriksson and Öckert (2014) estimate

the effect of tracking using a Swedish compulsory school reform as a natural experiment that changed a selective system with early tracking to a comprehensive system. They find that the month-of-birth effects are more pronounced by early tracking. Our contribution to the literature is to offer an alternative explanation for why month-of-birth effects persist by showing how parents, teachers, and peers influence the skill formation of children.

B Details of The Data

B.1 Supplementary Administrative Data

An anonymous municipality in the province provided administrative data on family SES and the names of high schools that students applied for and entered after grade nine. To examine the external validity of our analysis relying on the administrative data from the municipality, we compare the municipality and province in terms of key variables in our main data set. Table 8 summarizes the differences between the two for ninth graders. We omit the results for students in other grades because they do not differ substantially from those for grade nine. Cognitive and noncognitive skills as well as interpersonal relationships measures are normalized to have a mean of zero and a standard deviation of one for the whole province for the ninth graders.

Students in the municipality have about 0.2 standard deviations higher test scores than the provincial average, but lower self-control skills and poorer interpersonal relationships. In terms of the number of books at home, which is a proxy for family SES, the municipality is not significantly different from the provincial average. While the prep school attendance rate is slightly higher than the average, we find very little difference in time use.

Table 8: Difference Between the Province and Municipality

Subject	The province (A)	The municipality (B)	Difference (A-B)	Adjusted p-values
Japanese	0.000	0.187	-0.187	0.000
Math	0.000	0.223	-0.223	0.000
English	0.000	0.228	-0.228	0.000
Self-control	0.000	-0.298	0.298	0.000
Self-efficacy	0.000	-0.029	0.029	0.446
Conscientiousness	0.000	-0.053	0.053	0.129
Relative age	5.563	5.535	0.029	0.644
Girl	0.492	0.487	0.005	0.633
# of books: 0 ~ 10	0.135	0.124	0.011	0.094
Weekly hours of studying outside school	10.624	10.879	-0.255	0.062
Prep school	0.699	0.724	-0.024	0.007
Weekly hours of reading	4.060	3.843	0.216	0.220
Weekly hours of arts, music, and sports	1.128	1.115	0.013	0.840
Weekly hours of playing outside and sports	4.848	4.792	0.055	0.637
Peers	0.000	-0.054	0.054	0.002
Teachers	0.000	-0.190	0.190	0.000

Source: Provincial Skill Assessment 2015-2018.

Note: We compare the means of the key variables for the ninth grade students in the Skill Assessment between the municipality that provided administrative data and the province that conducted the Provincial Skill Assessment. Cognitive and noncognitive skills as well as interpersonal relationships measures are normalized to have a mean of zero and a standard deviation of one for the whole province for the ninth graders. The p-values are adjusted by a method that controls for a false discovery rate (see Benjamini and Hochberg (1995)) because there are 16 hypotheses.

B.2 Questions Measuring Noncognitive Skills

Table 9 lists questions used to measure noncognitive skills. The original questions are written in Japanese and translated into English by the authors.

Table 9: Questionnaire on Noncognitive Ability

Group	Question items
Self-control	I forgot something I needed for school.
	I interrupted other people.
	I said something rude.
	I could not find something because of the mess.
	I lost my temper.
	I did not remember what someone told me to do.
	My mind wandered.
Self-efficacy	I talked back when upset.
	I believe I will receive an excellent grade in this class.
	I'm certain I can understand the most difficult material presented in the readings for this course.
	I'm confident I can understand the basic concepts taught in this course.
	I'm confident I can understand the most complex material presented by the instructor in this course.
	I'm confident I can do an excellent job on the assignments and tests in this course.
Conscientiousness	I expect to do well in this class.
	I'm certain I can master the skills being taught in this class.
	Considering the difficulty of this course, the teacher, and my skills, I think I will do well in this class.
	I do my job without carelessness and inattention.
	I work hard and with pleasure.
	I engage myself in the things I do.
	During class time, I concentrate on the tasks given to me.
	When I finish my homework, I check it many times to see if I did it correctly.
	I respect the rules and help maintain order.
	If I make an appointment, I keep it.
	My room is in order.
When I start to do something I have to finish it at all costs.	
I like to keep all my school things in good order.	
I play only when I have finished my homework.	
It is unlikely that my attention will be diverted.	
I take care of my responsibilities.	

B.3 Correlations Between Cognitive and Noncognitive Skills

Table 10 reports the correlation coefficients between test scores and noncognitive skills. All three types of noncognitive skills are positively and significantly correlated with test scores in grades six and nine. The correlation with conscientiousness is about 0.15-0.20. The correlation with self-control is slightly weaker, at 0.10-0.20, and decreases with grade. The correlation with self-efficacy is much stronger than that for the other two noncognitive skills and becomes increasingly strong with

grade. The correlation coefficients are in the range of 0.30-0.50. Of course, correlation by no means indicates causation, but does demonstrate that it is useful in predicting cognitive skills.

Table 10: Correlation Coefficients Between Cognitive and Noncognitive Skills

		Conscientiousness	Self-control	Self-efficacy
		(1)	(1)	(2)
Grade 6	Math	0.171 (0.005)	0.140 (0.005)	0.375 (0.004)
	Japanese	0.181 (0.005)	0.139 (0.005)	0.260 (0.005)
Grade 9	Math	0.172 (0.005)	0.103 (0.005)	0.478 (0.004)
	Japanese	0.152 (0.005)	0.104 (0.005)	0.372 (0.004)
	English	0.196 (0.005)	0.098 (0.005)	0.450 (0.004)

Source: Provincial Skill Assessment 2016-2018. Standard errors are in parentheses.

C Additional Analysis

C.1 Month-of-Birth Gaps in Skill Growth

A potential problem with the regression analysis is that the composition of students differs across grades because some students move in or out of the province or switch between private and public schools. While the annual attrition rate is low (see Section 3.1), some of the observed convergence in the test score gaps might be driven by changes in the composition of the individuals in the tests. To address this concern, taking advantage of the panel structure of the data, we regress the growth (i.e., 1-year difference) as opposed to the level of test scores on age and other control variables. Table 11 shows the estimated gap between the oldest and youngest students (see Equation 2). The growth in test scores is faster for the youngest than for the oldest children in most grades, which implies that month-of-birth gaps in

the test scores diminish over time. Note that all estimates in columns (1)-(3) of Table 11 are statistically significant at the 1% level when p-values are adjusted for multiple testing of 11 hypotheses.

We again address biases arising from possible changes of student composition. We regress the growth of noncognitive skills on age and other control variables. Table 11 indicates that there is no significant gap in the growth of noncognitive skills in all grades, which implies that month-of-birth gaps in noncognitive skills do not converge over time. Again, these estimates are precisely estimated.

Table 11: Month-of-Birth Gaps in the Growth of Test Scores and Noncognitive Skills

	Japanese	Math	English	Conscientiousness	Self-control	Self-efficacy
	(1)	(2)	(3)	(4)	(5)	(6)
G4 to G5	-0.097 (0.009)	-0.080 (0.009)			-0.010 (0.015)	
G5 to G6	-0.114 (0.009)	-0.055 (0.009)			-0.009 (0.014)	-0.014 (0.015)
G6 to G7	-0.063 (0.009)	-0.016 (0.009)		-0.011 (0.015)		-0.010 (0.015)
G7 to G8	-0.095 (0.010)	-0.085 (0.008)		-0.006 (0.015)	-0.003 (0.015)	
G8 to G9	-0.010 (0.008)	-0.062 (0.009)	-0.052 (0.008)		0.008 (0.013)	0.020 (0.015)
Control variables	✓	✓	✓	✓	✓	✓

Note: The gaps in 1-year growth of an outcome between the oldest and youngest students are reported. Control variables include school fixed effects, survey year fixed effects, gender, and number of books at home. Standard errors are in parentheses. All estimates in columns (1)-(3) are statistically significant at the 1% level when p-values are adjusted for multiple testing of 11 hypotheses using a method that controls for a false discovery rate (see Benjamini and Hochberg (1995)). None of the estimates in columns (4)-(6) are statistically significant at any conventional level when p-values are adjusted.

C.2 Month-of-Birth Effects on High School Quality

One might argue that students may not take the academic tests and questionnaire for noncognitive skills seriously because there is no incentive to do so. To address this concern, we estimate month-of-birth effects on a high-stake academic outcome, which is the quality of high school that students enter after grade nine. High school education is not compulsory in Japan, but 98% of students enter high school immediately after the ninth grade. Students can apply for any public high school in the province and any private high school in/outside the province. Oversubscribed schools select students based on a written examination and grade point average. Very few schools interview candidates. The quality of high school in Japan is often measured by academic test scores, which are perceived as a requirement for entrance.

A municipality in an anonymous province provided us with information on the high school their students entered after grade nine. We link these data with the main skill assessment data and the quality of high school published by a private firm.¹² The measured quality was originally scaled to a mean of 50 and a standard deviation of 10. The sample size is 937, as the data cover only one municipality in the province.

Figure 8 shows clearly that high school quality increases with month of birth. We also regress high school quality on age and other control variables using equation 1. Table 12 presents the estimated gap in high school quality between the oldest and youngest students among the ninth graders (see equation 2). Without the control variables, the oldest students enter 4.653 points higher quality high schools than the youngest students in the same class (see column 1). Controlling for school fixed effects, family SES, and gender does not change the estimate (see column 2). Additionally controlling for math, Japanese, and English test scores conducted at the beginning of the ninth grade reduces the gap substantially, but the remaining gap is still sizable and statistically significant (see column 3). The estimate implies

¹²The data were downloaded from minkou.jp in October 2018. A marketing research company claims that this website is the most viewed among those providing information on schools in Japan. Other firms also publish high school quality data, but they are similar to each other. We also confirm that teachers, students, and their parents often referred to the website during informal interviews.

that about two-thirds of the month-of-birth gaps in high school quality are explained by cognitive skills. Finally, we control for self-efficacy as a noncognitive skill in addition to all other control variables, but the estimate in column 4 does not change relative to that without controlling self-efficacy. We also note that the results do not change when we use the quality of high school for which students applied instead of the high school they entered because they match most of the cases.

To put the estimates in context, we also report estimated test score gaps for the ninth graders in this municipality. The estimates indicate that the oldest students have 0.381, 0.329, and 0.253 standard deviations higher scores than the youngest students in math, Japanese, and English, respectively. These gaps are about twice as large as those for the whole province (see Table 4), which suggests that the high school quality gap may also be substantially greater than that for other municipalities in the province.

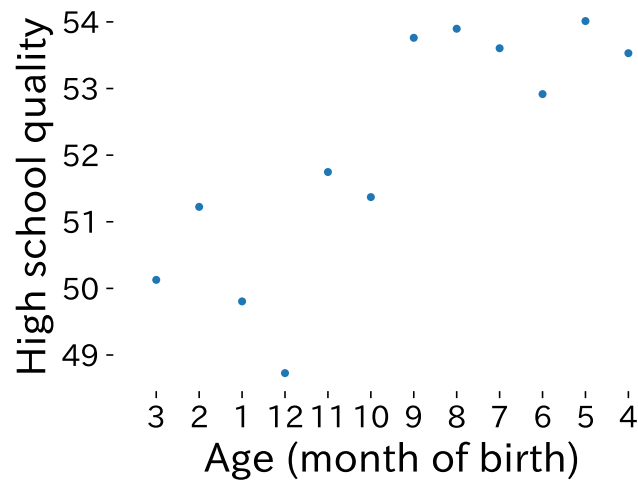


Figure 8: Average High School Quality by Month of Birth

Source: Administrative data in 2017 from a municipality in the province.

Note: High school quality is normalized so that the mean is 50 and the standard deviation is 10.

The number of observations is 937.

Table 12: Month-of-Birth Effects on High School Quality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent var.		High school quality			Math	Japanese	English
Month of birth effect	4.653 (1.110)	4.521 (1.040)	1.422 (0.699)	1.327 (0.688)	0.381 (0.099)	0.329 (0.099)	0.253 (0.102)
School FE, SES, gender		✓	✓	✓	✓	✓	✓
Test scores			✓	✓			
Self-efficacy				✓			
No. of obs.	937	931	930	916	961	961	961

Source: Administrative data from a municipality in the province that conducted the skill assessment.

Note: The estimated differences between the oldest (April-born) and youngest (March-born) students in the same grade are reported. High school quality is normalized so that the mean is 50 and the standard deviation is 10. The standard errors are in parentheses.

C.3 Heterogeneity of Month-of-Birth Effects

C.3.1 By Gender and SES

The literature shows that boys mature more slowly than girls and are more susceptible to negative family and school environments, which suggests that month-of-birth effects may be stronger for boys than for girls. Similarly, children from low-SES families may be more susceptible to month-of-birth effects because they have fewer resources to fund compensatory behavior. Although some papers find that month-of-birth effects tend to be stronger for boys and/or low-SES children, the overall evidence is mixed.¹³ We provide an additional piece of evidence about the existence of heterogeneous effects by gender and SES.

Table 13 reports the estimated girl vs. boy gaps in the month-of-birth effects, i.e., the gap between ΔY for girls and that for boys (see equation (2)). None of the estimates are statistically significant when we adjust p-values for multiple testing with 29 (6 + 6 + 2 + 4 + 6 + 5) hypotheses. The results do not change when we pool observations across grades as long as the p-values are adjusted for six hypotheses. We conclude that there is no evidence in our data of gender heterogeneity in month-of-birth effects on cognitive and noncognitive skills.

Table 14 reports the estimated gap in month-of-birth effects ΔY between low-SES and other students. Our proxy for being a low-SES student is having 0-10 books at home. Again, none of the estimates are statistically significant when p-values are adjusted for multiple testing. When we pool observations across grades, only the estimated effect on math is statistically significant, even if the p-values are adjusted. The estimate implies that the month-of-birth effect for low-SES students is 0.033 standard deviations smaller than that for other students.

Table 15 shows estimates using an alternative measure of low SES. Here, we use administrative data from the municipality in the province and define low SES by whether the student's family receives an education subsidy for low-income house-

¹³Puhani and Weber (2007), McEwan and Shapiro (2008), and Matta et al. (2016) find stronger month-of-birth effects for boys than for girls, while Peña (2017) and Attar and Cohen-Zada (2018) find that the gender gap is insignificant. Only a few papers examine heterogeneity by family SES. Matta et al. (2016) find that the effects are stronger for low-SES children, while Attar and Cohen-Zada (2018) find no heterogeneity by family SES.

holds. We find that the month-of-birth effects on self-efficacy is significantly stronger for low-income households in this subsample. We find no evidence for heterogeneous treatment effects on other outcomes.

Table 13: Heterogeneous Month-of-Birth Effects by Gender

	Math (1)	Japanese (2)	English (3)	Conscientiousness (4)	Self-control (5)	Self-efficacy (6)
Grade 4	-0.008 (0.017)	0.011 (0.016)			-0.017 (0.030)	
Grade 5	0.005 (0.016)	0.013 (0.016)			0.007 (0.031)	0.051 (0.029)
Grade 6	0.008 (0.017)	0.008 (0.017)		0.040 (0.030)	0.004 (0.030)	0.028 (0.030)
Grade 7	0.004 (0.017)	-0.018 (0.018)		0.059 (0.027)	-0.007 (0.030)	0.045 (0.031)
Grade 8	-0.003 (0.017)	-0.008 (0.019)	-0.002 (0.017)	0.011 (0.029)	-0.005 (0.031)	-0.003 (0.028)
Grade 9	-0.002 (0.017)	-0.004 (0.017)	-0.011 (0.017)	-0.041 (0.030)	-0.020 (0.031)	0.031 (0.028)
All	0.000 (0.009)	0.001 (0.009)	-0.006 (0.015)	0.015 (0.016)	-0.006 (0.018)	0.030 (0.016)
Controls	✓	✓	✓	✓	✓	✓

Source: Provincial Skill Assessment 2015-2018.

Note: The gaps in month-of-birth effects (based on equation 2) between boys and girls are reported. The control variables include school fixed effects, survey year fixed effects, gender, and number of books at home. Standard errors are in parentheses. None of the estimates are statistically significant when p-values are adjusted for multiple testing, using a method that controls for a false discovery rate (see Benjamini and Hochberg (1995)).

C.3.2 Quantile Regression

Du et al. (2012), Muller and Page (2016), and Tukiainen et al. (2017) find that month of birth affects the probability of occupying top management and political positions, which suggests that month-of-birth effects might be stronger at the right tail of an ability distribution. To address this issue, we estimate month-of-birth effects on test scores using quantile regressions. Unfortunately, quantile regression

Table 14: Heterogeneous Month-of-Birth Effects by SES (No. of Books at Home)

	Math (1)	Japanese (2)	English (3)	Conscientiousness (4)	Self-control (5)	Self-efficacy (6)
Grade 4	-0.029 (0.023)	-0.001 (0.026)			0.061 (0.053)	
Grade 5	-0.005 (0.030)	-0.044 (0.030)			-0.046 (0.054)	-0.033 (0.057)
Grade 6	-0.023 (0.030)	0.008 (0.030)		-0.097 (0.063)	0.082 (0.059)	-0.050 (0.060)
Grade 7	-0.036 (0.028)	-0.025 (0.026)		-0.060 (0.049)	-0.010 (0.051)	-0.059 (0.049)
Grade 8	-0.011 (0.027)	0.001 (0.026)	-0.014 (0.026)	0.107 (0.053)	0.008 (0.047)	-0.067 (0.045)
Grade 9	-0.060 (0.027)	-0.052 (0.026)	-0.048 (0.024)	0.041 (0.045)	-0.034 (0.049)	-0.023 (0.044)
All	-0.035 (0.013)	-0.026 (0.013)	-0.032 (0.020)	0.001 (0.027)	0.006 (0.024)	-0.046 (0.025)
Controls	✓	✓	✓	✓	✓	✓

Source: Provincial Skill Assessment 2015-2018.

Note: The gaps in month-of-birth effects (based on equation 2) between low-SES students and others are reported. We define low-SES students as those who have 0-10 books at home. The control variables include school fixed effects, survey year fixed effects, gender, and number of books at home. Standard errors are in parentheses. All but one of the estimates are statistically significant when p-values are adjusted for multiple testing, using a method that controls for a false discovery rate (see Benjamini and Hochberg (1995)). The exception is the estimate for math when observations are pooled across grades (labeled as All).

Table 15: Heterogeneous Month-of-Birth Effects by SES (Family Income)

	Math (1)	Japanese (2)	English (3)	Conscientiousness (4)	Self-control (5)	Self-efficacy (6)
All	0.095 (0.069)	0.088 (0.067)	-0.100 (0.153)	0.049 (0.081)	-0.049 (0.229)	0.346 (0.097)
Adjusted p-values	0.376	0.376	0.651	0.651	0.829	0.002
No. of obs	11698	11695	2985	1001	2066	988
Controls	✓	✓	✓	✓	✓	✓

Source: Provincial Skill Assessment and administrative data from a municipality in 2015 and 2018. Note: The gaps in month-of-birth effects (based on equation 2) between low-SES students and others are reported. We define low-SES students by whether their family receives an education subsidy for low-income households. The control variables include school fixed effects, survey year fixed effects, gender, and number of books at home. Standard errors are in parentheses.

for noncognitive skills turns out to be uninformative because it takes only a few distinct values, and hence, many estimates are exactly zero.

Table 16 presents the estimated gaps between the oldest and youngest students in the same cohort for the 10th, 50th, and 90th percentiles. We pool primary school students (grades 4-6) and junior high school students (grades 7-9) to reduce noise. For math and Japanese test scores, month-of-birth effects are more pronounced toward the right tail of the distribution. While the gaps between percentiles are statistically significant (see Table 17), the magnitude does not seem very large. For example, the month-of-birth gaps in math score between the oldest and youngest junior high school students for the 10th and 90th percentiles are 0.151 and 0.202, respectively.

C.4 Comparison of Different Grades for Same Age

We also estimate cognitive and noncognitive skill gaps between students born before and after the cutoff date. Because all students take the tests in the same month, the former or March-born students have been in school a year longer than the latter or April-born students, while both are nearly the same absolute age. The gaps are estimated using the regression discontinuity design (RDD), and the econometric model is given by:

Table 16: Month-of-Birth Effects for 10th, 50th, and 90th Percentiles

		10 th	50 th	90 th
Japanese	G4-G6	0.284 (0.007)	0.297 (0.005)	0.324 (0.004)
	G7-G9	0.202 (0.009)	0.222 (0.006)	0.230 (0.004)
Math	G4-G6	0.274 (0.007)	0.258 (0.005)	0.352 (0.010)
	G7-G9	0.151 (0.007)	0.164 (0.005)	0.202 (0.008)
English	G4-G6			
	G7-G9	0.098 (0.009)	0.133 (0.007)	0.046 (0.005)

Table 17: 10th-50th and 50th-90th Percentile Gaps in Month-of-Birth Effects

		10th - 50th	50th - 90th
Japanese	G4-G6	-0.013 (0.007) (-0.033 -0.005)	-0.027 (0.010) (-0.053 -0.013)
	G7-G9	-0.020 (0.011) (-0.029 0.016)	-0.007 (0.009) (-0.037 -0.001)
Math	G4-G6	0.016 (0.008) (0.001 0.034)	-0.094 (0.009) (-0.110 -0.077)
	G7-G9	-0.013 (0.008) (-0.021 0.007)	-0.038 (0.013) (-0.053 -0.011)
English	G7-G9	-0.035 (0.014) (-0.070 -0.017)	0.088 (0.013) (0.066 0.117)

Note: Month-of-birth effects are estimated using quantile regression. All models include school fixed effects, survey year fixed effects, gender, and number of books at home. The standard errors for the gaps in the estimates are based on nonparametric bootstrap with 100 replications.

$$Y_{it} = \gamma_0 + \gamma_1 M_{it} + \gamma_2 M_{it}^2 + \gamma_3 D_{it} + \gamma_4 D_{it} M_{it} + \gamma_5 D_{it} M_{it}^2 + \gamma_6 X_{it} + v_{it}, \quad (3)$$

where Y_{it} is an outcome of student i in year t , M_{it} represents the distance from the cutoff measured in months, D_{it} is an indicator that equals one if student i was born before the cutoff date and zero otherwise, X_{it} is a vector of control variables that include gender and year and school fixed effects, and v_{it} is an error term.

The parameter of interest here is γ_3 , which reflects three different factors. First, this parameter captures the effect of 1-year schooling. Students born before the cutoff have been in school a year longer than those born after the cutoff. Second, it captures the effect of school starting age. Students born before the cutoff started schooling a year earlier. We cannot identify these two effects separately because of perfect collinearity. Third, the parameter reflects relative age effects. Students born before the cutoff are the youngest in their grade, while those born after the cutoff are the oldest in their grade. It is hard to disentangle the relative age effect from other factors because there is little variation in relative age.

The key identifying assumption for this RDD model is that students' characteristics other than birth month must be continuous at the school-entry cutoff. This assumption is violated if parents manipulate children's birth month, which is the same as the identifying assumption for the regression model (1).

Figure 9 shows graphically the estimated test score gaps between two adjacent grades for students of the same age. The estimates, or the coefficients for grade dummy (γ_3) in equation (3), are also reported in Table 18. For the math test, the gap between grades four and five is 0.078, the gaps are insignificant for grades five to eight, and the gap is 0.233 between grades eight and nine. For the Japanese test, the grade gaps are small in magnitude except for the gap between grades eight and nine. Finally, for the English test, the gap between grades eight and nine is 0.537, which is greater than the gaps for the math and Japanese tests.

For all subjects, the gap between grades eight and nine is largest, presumably because ninth graders work more hours per week and are more likely to go to prep school to get ready for high school entrance exams. To be specific, 70% of ninth

graders go to prep school, compared with 60% of eighth graders. Ninth graders study 10.6 hours per week outside of school, compared with 9.6 hours for eighth graders. The gap in the English test is greater than that for the math and Japanese tests, which reflects the gap in accumulated years of studying the subjects. Eighth graders have studied English for only a year, which is half the accumulated years of studying English for ninth graders. By contrast, eighth graders have studied math and Japanese for 7 years, which is seven-eighths of the accumulated years of studying for ninth graders.

Table 18: Test Score Gaps Between Grades for the Same Age

	Math (1)	Japanese (2)	English (3)
G4 vs. G5	0.078 (0.009)	0.005 (0.013)	
G5 vs. G6	-0.013 (0.009)	0.030 (0.011)	
G6 vs. G7	-0.016 (0.009)	0.076 (0.010)	
G7 vs. G8	0.015 (0.009)	0.033 (0.009)	
G8 vs. G9	0.233 (0.009)	0.189 (0.010)	0.537 (0.009)

Source: Provincial Skill Assessment 2015-2018.

Note: Test score gaps for students in different grades but of the same age are reported using equation (3). The controls include school fixed effects, survey year fixed effects, gender, and number of books at home. Standard errors are in parentheses.

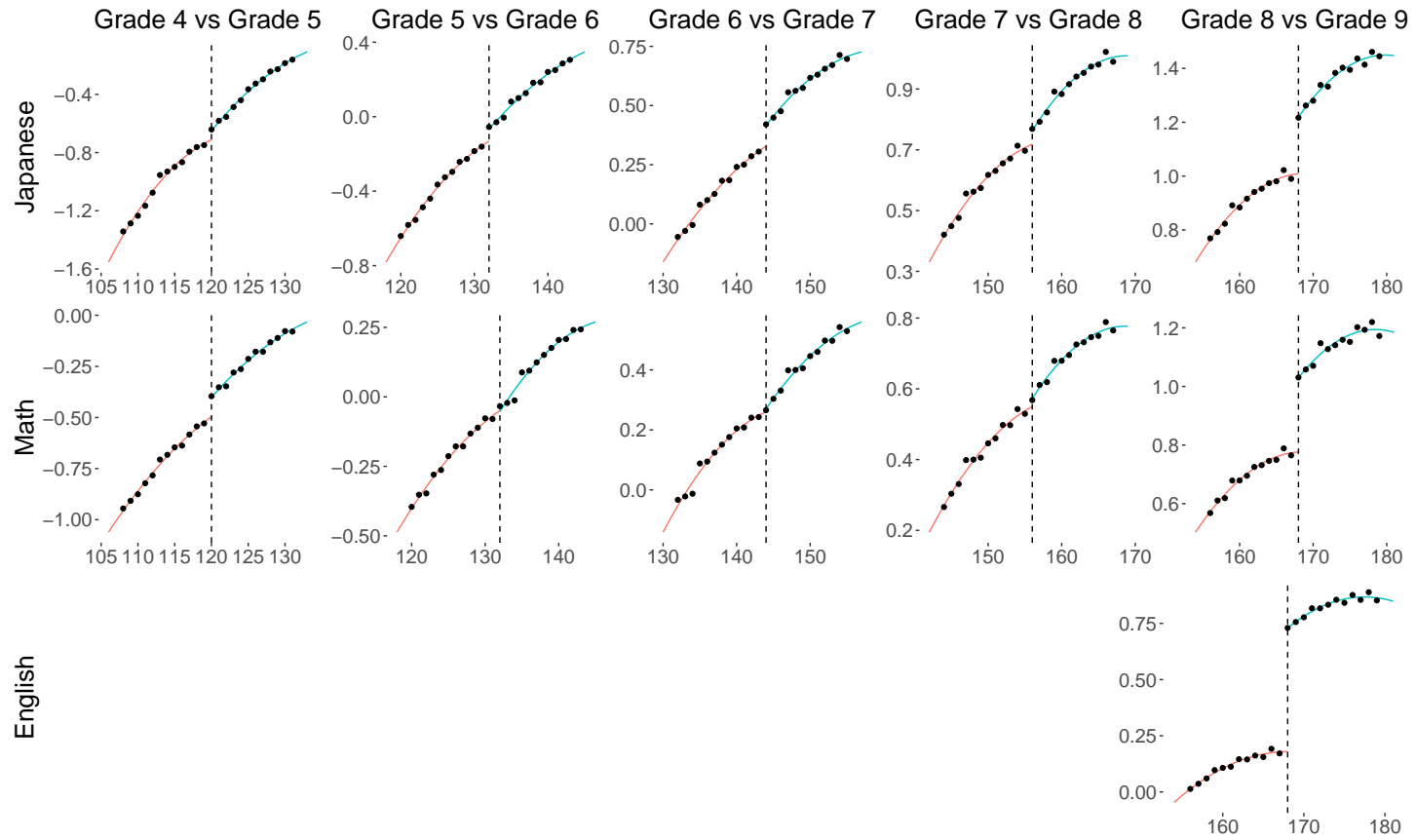


Figure 9: Test Score Gaps Between Grades for the Same Age

Source: Provincial Skill Assessment 2015-2018.

C.5 Correlation Between Time Use and Interpersonal Relationships

We estimate how time use and interpersonal relationships are associated with cognitive and noncognitive skills. We regress these skills on hours of study, prep school attendance, relationships with teachers, and relationships with peers along with other control variables, including gender, number of books at home, and school and year fixed effects. In addition, we also estimate models with student fixed effects. Other time use variables, such as hours of reading, are available only for 2015, and hence, cannot be included in a fixed-effect regression. All models are estimated separately for elementary and junior high schools and by subject.

The regression results for test scores are summarized in Table 19. We focus on results with student fixed effects, but OLS results are also reported for reference. For both math and Japanese and for both elementary and junior high school, all the educational inputs and interpersonal relationships are positively correlated with test scores given student fixed effects. Table 19 also shows the results for regressions of noncognitive skills. Hours of study, relationship with teachers, and relationship with peers are positively associated with noncognitive skills, whereas only prep school attendance is negatively associated with these three noncognitive skills. These results are at best suggestive and should be carefully interpreted because they by no means imply causality.

Table 19: Regressions of Cognitive and Noncognitive Skills on Inputs and Personal Relationships

	Grade 4-6				Grade 7-9			
	Math		Japanese		Math		Japanese	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Hours of studying	0.025 (0.001)	0.006 (0.001)	0.024 (0.001)	0.004 (0.000)	0.026 (0.000)	0.009 (0.000)	0.023 (0.000)	0.005 (0.000)
Prep school	-0.001 (0.004)	0.024 (0.005)	-0.086 (0.004)	0.021 (0.005)	0.163 (0.006)	0.081 (0.005)	-0.039 (0.006)	0.004 (0.005)
Teachers	0.012 (0.002)	0.007 (0.003)	0.025 (0.002)	0.007 (0.002)	-0.009 (0.002)	0.008 (0.002)	0.001 (0.002)	0.006 (0.003)
Peers	0.078 (0.002)	0.008 (0.003)	0.102 (0.002)	0.004 (0.002)	0.074 (0.002)	0.006 (0.002)	0.088 (0.002)	0.004 (0.002)
Control variables	✓	✓	✓	✓	✓	✓	✓	✓
Individual fixed effect		✓		✓		✓		✓

	Conscientiousness		Self-control		Self-efficacy	
	(1)	(2)	(3)	(4)	(5)	(6)
Hours of studying	0.038 (0.001)	0.018 (0.001)	0.027 (0.000)	0.011 (0.000)	0.043 (0.000)	0.018 (0.001)
Prep school	-0.095 (0.009)	-0.020 (0.008)	-0.078 (0.005)	-0.023 (0.006)	0.052 (0.005)	0.019 (0.006)
Teachers	0.106 (0.006)	0.058 (0.004)	0.141 (0.002)	0.070 (0.003)	0.066 (0.003)	0.051 (0.003)
Peers	0.180 (0.006)	0.067 (0.004)	0.195 (0.002)	0.067 (0.003)	0.190 (0.003)	0.076 (0.003)
Control variables		✓	✓	✓	✓	✓
Individual fixed effect		✓		✓		✓

Source: Provincial Skill Assessment 2015-2018.

Note: Control variables include gender, number of books at home, and school and year fixed effects. Standard errors are in parentheses.