I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14077

Adolescence Development and the Math Gender Gap

Cristina Borra
Maria Iacovou
Almudena Sevilla

JANUARY 2021

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14077

Adolescence Development and the Math Gender Gap

Cristina Borra
Universidad de Sevilla and IZA

Maria Iacovou
University of Cambridge

Almudena Sevilla
UCL and IZA

JANUARY 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Adolescence Development and the Math Gender Gap

Using different production function models, we study the causal association between adolescence development and the increase in the gap in math performance between boys and girls. We use data from the 1958 British National Child Development Study, a longitudinal survey of all British children born in the first week of March 1958, containing unique information on puberty development and educational outcomes from childhood into adolescence. We first document a widening of about 10 percent of a standard deviation in the gender gap in maths from primary to secondary school in the UK, and show that adolescent development contributes to explain almost two thirds of the widening of the math gender gap during the adolescence years. We also explore the mechanisms behind these effects. Our evidence regarding differences in the impact of puberty development by age, subject and self-perceived math ability suggests that both social conditions and biological factors are behind the estimated relationships between adolescent development and the increase in the gender gap in math in secondary school.

JEL Classification: I21, I24, J16

Keywords: pubertal development, educational outcomes, gender gap in mathematics

Corresponding author:
Cristina Borra
University of Seville
Calle San Fernando, 4
41004 Sevilla
Spain
E-mail: cborra@us.es

1. Introduction

Most countries have witnessed a large increase in female human capital accumulation and labour force participation. Gender differences in academic attainment and achievement have dramatically reversed in the last decades in the United States and whereas in the 1960s there were 1.6 males for every female graduating from four-year colleges, there are now 1.35 females for every male (Goldin et al. 2006). Similarly, across OECD countries while in 1998 21% of adult men had a college degree vs. 19% of adult women, by 2018 the relative shares had reversed with 40% of women holding a college degree vs. 34 % of men (OECD, 2019a). However, the literature has consistently documented a higher academic achievement in mathematics of boys over girls in many countries (Guiso et al 2008, Kane and Mertz 2012). On average girls' math results are 9% of a standard deviation lower than boys' at 15 years of age, although there is substantial variation across countries (see for instance Borgonovi et al. 2018 using data on 13 OECD countries) and the gender gap appears to be larger at the top of the ability distribution (Machin and Pekkarinen 2008, Ellison and Swanson 2010). This paper explores whether adolescence development can explain the different trends over the lifecycle in math performance by boys and girls, and the mechanisms underlying this effect.

A better understanding of the mechanisms that facilitate mathematics performance of boys and girls is essential from the point of view of equal opportunities policies and gender equality. Despite the remarkable narrowing differences in educational outcomes, women continue to be a very low proportion of all graduates in science, technology, engineering and mathematics (STEM). In the United States, Kahn and Ginther (2017) show that in 2014 women received only 27% of undergraduate degrees in the math-intensive STEM fields, compared to 69% of undergraduate degrees in other STEM fields, including social sciences. Across the developed world only 31% of women graduate in STEM (OECD, 2018). This underrepresentation of women in math intensive STEM fields has important consequences for gender inequality. Math skills during school years appear to constitute a good predictor of readiness for STEM programs at university, which offer higher wages (Black et al. 2008; Blau and Kahn 2017). For instance, students completing at least 3 math and science classes in the last year of high school have an 80% chance of enrolling in STEM programs at university, while those failing to meet this standard have just a 5% chance of entering any STEM program (Card and Payne 2017). Mathematics is a valuable skill in and of itself. An additional course of study in algebra and geometry in high school increases future earnings between

2.5 and 3.2 percent (Rose and Betts 2004). And students induced to take advanced mathematics courses earn about 3 percent more in Denmark (Joensen and Nielsen 2015).

A series of studies have recently documented a widening of the gender gap in math from primary into secondary school in Chile, Italy, and the US (Bharadwaj et al. 2016; Contini et al 2017; Ellison and Swanson 2018). On average the math gender gap more than triples from about 3% of a standard deviation at age 9/10 to about 9% at age 15/16 (Borgonovi et al. 2018). Using different cohort-based data sources for the UK spanning over three decades, we also document a widening of the math gender gap during the adolescence period in the UK that is persistent over time. In particular, the math gender gap increases from 0.08 standard deviations at 7 years of age to 0.18 at 16 years old for the 1958 cohort of the National Child Development Study (NCDS58). For the 1990 cohort of the Avon Longitudinal Study of Parents and Children (ALSPAC90), the math gender gap increases from a null figure at age 7 to 0.12 standard deviations at age 14. Similarly, we report gender gaps in maths over time for later cohorts using available data for England from TIMMS and PISA (Mullis et al. 2016; OECD, 2010,2014,2019b). We are able to identify gaps for the 1993 cohort, interviewed at age 9-10 in 2003 TIMMS and at age 15-16 in 2009 PISA; for the 1996-1997 cohort, interviewed at age 9-10 in 2007 TIMMS and at age 15-16 in 2012 PISA; and for the 2001-2002 cohort, interviewed at age 9-10 in 2011 TIMMS and at age 15-16 in 2018 PISA. For all cohorts, from non-significant differences in math achievement at age 9, we reach significant positive gaps in favour of boys at age 15-16. The size of the gap seems to have diminished over time from about 0.21 standard deviations in 2009 to a gap of about 0.12 standard deviations in 2012 and 2018, but it still remains significant.

The mechanisms behind the increase in the gender gap in mathematics from childhood to adolescence remain poorly understood. Recent research in neuroscience recognizes now that adolescents' brains are physically different from younger children's brains and adults' brains (Blakemore 2018 p.73). Adolescence is a sensitive period, second to early childhood, when connections between neurons (synapses) can be pruned or strengthened to increase neural efficiency (Blakemore and Choudhury 2006, Steinberg 2014, Dahl et al. 2018).[1] The economics literature has also shown a

[1] In particular during adolescence grey matter diminishes in volume as synapses are pruned, while white matter increases as axons are myelinated to increase information speed (Giedd et al. 1999, Blakemore et al 2010). These processes are accompanied by higher efficiency in cognitive function including improvements in intelligence quotient, working memory, problem solving, executive functioning, and social cognition (Steinberg et al. 2009, Fuhrmann et al. 2015). Longitudinal studies suggest that the physical and hormonal changes during puberty may directly influence the grey matter decreases and white matter increases taking place during adolescence (Herting et al., 2015; Goddings et al. 2014; Piekarsky et al.

renewed interest in the adolescence period from studying risky behaviour (Aizer 2017) to the way adolescence children invest their time (del Boca et al. 2017) to non-cognitive skills (Lundberg 2017 and Schurer 2017). It has documented that gender differences in competitiveness, leadership, and risk preferences kick in around puberty. For instance, Alan et al (2019) show that, while there is no gender difference in the willingness to become a leader among young children, a large gap emerges around puberty. Andersen et al (2013) document a similar finding for competitiveness. Also, Andreoni et al. (2019) find that while there are no differences in risk taking among children aged 3 to 12, 13-to-15-year-old girls are significantly more risk averse than similarly aged boys, suggesting that the gap in risk taking emerges in adolescence.

The fact that puberty encompasses physical and hormonal changes opens the door for biology to affect cognitive abilities and behaviour, and the fact that boys and girls reach adolescence at different points in their life cycle may also contribute to an emergence of a math gender gap. It is also possible that the physical changes taking place during adolescence may also affect the expected behaviour of boys and girls. Asynchronous development may for instance cause problems in psychosocial adaptation increasing the chances of risky behaviours (Brooks-Gunn et al 1985, Haynie 2003). Also, social norms to develop stereotypical gender identities may become stronger with puberty (Akerlof and Kranton 2000). In fact, Andersen et al (2013) compare competitiveness differences in a matrilineal and a patriarchal society and show that whereas there are no differences in competitiveness between boys and girls at any age in the matrilineal society, girls become less competitive around puberty in the patriarchal society. They conclude that culture and socialization appear to interact with biological forces in determining the difference in gender gaps between the matrilineal and patriarchal societies around puberty. This view is also shared by neuroscientists that believe that from puberty onwards the social world around us also influences brain development, that is, which neural connections are pruned, and which are retained (Blakemore 2018, Dahl et al. 2019).

We use a unique dataset, the first four waves (pregnancy and birth, and 7, 11, and 16 years-old) of the NCDS58. The NCDS is specially indicated to study the impact of puberty development on student outcomes because it is the only database to offer, besides self-reported assessments, medical assessments of puberty development alongside a rich set of cognitive outcomes. Medical assessments

2017). In consequence pubertal development may directly impact social and cognitive abilities. (Picci and Scherf 2016; Steinberg 2008). Moreover, pubertal development may have sex-specific effects on trajectories of brain maturation and neurodevelopment in boys and girls, which may give rise to important differences in social and cognitive abilities by gender (Brouwer et al. 2015).

of puberty were performed alongside the rest of medical examinations minimizing sample attrition, and are generally considered the gold standard in the evaluation of puberty development (Baird et al. 2017). Its longitudinal character allows implementing very rich dynamic production function models to estimate the causal effect of puberty, and the availability of a rich set of individual (endowments), household, and school characteristics allows us to control for many arguably exogenous variables that very rarely the literature can control for.

We first present novel causal evidence on the impact of puberty development on boys' and girls' math test scores at 11 and 16 years of age by estimating very rich production function models of boys' and girls' math skills, paying special attention to the assumptions needed to identify key parameters in each model (see Todd and Wolpin 2003). We estimate contemporaneous, value added, cumulative value-added models (see Fiorini and Keane 2014). We address two key challenges for identification. First, we control for measurement error in the lagged test score through an instrumental-variables model, using the twice-lagged skill as an instrument as in Del Bono et al. (2016). Second, we minimize selection bias due to unobservable factors by using a very rich dataset that allows controlling for a myriad of variables. We also perform falsification analyses showing that puberty development is unrelated to maternal education, maternal smoking during pregnancy, birth weight, and pre-term birth, suggesting that unobserved factors such as maternal IQ or nutrition patterns are not driving our results. We also show the robustness of our findings to departures from the cumulative value-added model proposed by Del Boca et al. (2017) and Agostinelli et al (2019) that consider additional information on reading scores and non-cognitive skills in estimating math skills production functions.

Our findings suggest that puberty development may explain the widening in the math gender gap over the lifecycle, especially its widening during the adolescence years. Using different models and puberty measures, we find that puberty development impacts differently boys' and girls' math outcomes, benefiting girls' outcomes for girls who show advanced development at 11 years of age but impairing the math performance of girls who show advanced development at 16 years of age, and improving -or not harming, depending on the specification- the math performance of boys that show advanced development at 16 years of age. We find that the different timing in and sign of the impacts of puberty on math scores by gender can explain about two thirds of the math gender gap. The math gender gap drops from about 20 percent to about 8 percent of a standard deviation at age 16 when puberty development is controlled for.

Further analysis into the mechanisms through which pubertal development may affect math outcomes suggests that social conditions interact with biological factors in shaping gender gaps in maths. First, if the channel was purely biological the impact of adolescent development should be similar independently of the child's age. We find that, at least for girls, puberty development is positive for math cognitive development at 11 years of age but negative later on. Second, a purely biological channel would also imply that adolescent development would impact math attainment in the same way independently of the respondent's self-assessed ability in math. Heterogenous analyses reveal that there is a much larger impact of puberty development on boys showing higher self-perceived assessment of math skills. If puberty development impacted math outcomes purely through a biological mechanism, we should expect to see the same effects from puberty on math and on English performance by boys and girls. We find that puberty development influences 16-year-old girls' English and math skills differently, as would be consistent with a social stereotype explanation of gender gaps in cognitive outcomes. These results are consistent with previous economic literature emphasizing the role of both self-confidence and social-stereotypes in explaining math gender gaps (Coffman 2014 and Bordalo et al. 2019) and prior child development literature documenting increases in self-esteem upon development for boys, and decreases for girls (Martin and Steinbeck 2017).

We contribute to the novel and growing literature documenting a persistence of the widening of the math gender gap over time by highlighting the causal role of puberty in widening the math gender gap upon puberty and the possible mechanisms at play. A limited number of studies has looked at the contemporaneous influence of puberty on educational outcomes, estimating mere correlations, including no controls for birth endowments or schooling variables (Koivusilta and Rimpela 2004, Cavanagh et al. 2007, Dreber et al 2011, Koerselmann and Pekkarinen 2017, Martin and Steinberg 2017). None of the existing studies looking at puberty and educational outcomes has focused on the impact of puberty development on math's skills. Our study also contributes to the literature exploring the potential channels of gender differences on math academic achievements. Most studies use cross-section data and cannot include child endowment proxies. Out of the few longitudinal studies documenting the widening of the math gender gap during adolescence (Fryer and Levitt 2010, Contini et al. 2017) only the paper by Bharadwaj et al (2016) includes birth characteristics. This study is the first to estimate the dynamic causal impacts of puberty development on math outcomes.

Our work also contributes to the very recent literature emphasizing differences in economic preferences of boys and girls emerging during adolescence such as self-efficacy, competitiveness, leadership, and risk taking (Andersen et al. 2013, Alan et al 2019, Andreoni et al. 2019), and how gender stereotypes interact with self-confidence in explaining differences in educational performance and occupational choices by boys and girls (Coffman 2014; Reuben, et al. 2017, Bordalo et al. 2019, Carlana 2019). Puberty is commonly proxied by age in these studies, which makes it harder to understand the mechanisms at play. The richness of information in our data related to the child's health, education, and school and family environments allows us to test for the mechanisms behind the effect of puberty on math outcomes and to understand the differential effect for boys and girls.

This paper is organized as follows. Section 2 describes the dataset. Section 3 presents the different empirical specifications of the math skills production function. Section 4 presents main estimation results together with specification and identification checks. Section 5 discusses potential mechanisms. Finally, Section 6 concludes.

2. Data

The 1958 National Child Development Study (NCDS) is a longitudinal study that sampled all children born in the first week of March 1958. Originally designed as a perinatal mortality study, the initial birth survey was followed by different sweeps carried out when the cohort members were 7, 11, 16, 23 and 33 years of age. More recently, the UCL Centre for Longitudinal Studies has carried out the four most recent sweeps, when cohort members turned 41, 46, 50, and 55 years old (in 2013). The studies between ages 7 and 16 contain information on cognitive outcomes, schooling, health, and family circumstances. In particular, the NCDS has special features that make it especially suitable for the study of puberty influences on educational achievement. First, thanks to its longitudinal nature we can construct different puberty measures for ages 11 and 16, coincident with cognitive achievement measures. Second, it offers information on the child's school environment, such as school size, whether schooling is gender-segregated, and the type of tracking policies followed. And third, the study offers both reported and, importantly, medical assessments of puberty development. This last feature constitutes the main advantage of the NCDS over other longitudinal studies that also include education and puberty measures, such as the Avon Longitudinal Study of Parents and Children (ALSPAC), the Millenium Cohort Study (MCS), the US Child Development Supplement, the National Longitudinal Study of Adolescent to Adult Health (Add Health), and the Longitudinal

Study of Australian Children. These studies offer parent or child self-assessed measures of puberty development, which are generally considered to be less accurate than objective assessments by doctors (Brooks-Gunn and Warren 1985; Baird et al. 2017). Also, given that puberty examinations were performed together with the rest of medical examinations in the NCDS, sample attrition was minimized. Additionally, compared to other longitudinal studies in the UK such as ALSPAC and MCS, in the NCDS all children were born in the same week, and thus issues of differences in chronological ages are minimized.

We select how many individuals for which we have information on test scores at age 7 and both test scores and puberty measures at ages 11 and 16. Columns 1 to 3 in Table B.1 in Appendix B show the effect of attrition and missing data on the distribution of some observables. We acknowledge that our sample may not be adequately representative of the non-white population. However, the small size of the coefficients for the rest of the independent variables suggest that selection on observables is quantitatively weak.

2.1 Test scores

Our main dependent variables are standardized math scores. The NCDS58 administered math tests at ages 7, 11, and 16. At age 7, 10 problems were read to the child and the score ranged from 0 to 10. At age 11, there were 40 questions (scores ranged 0 to 40), and at age 16, there were 31 questions (scores ranged from 0 to 31). (See examples of math problems in Appendix C). Children were also administered reading tests at the same ages. There were 30 reading questions at age 7, 35 questions at age 11, and 35 questions at age 16. (See examples of reading questions in Appendix C). All scores are standardized to have a mean of zero and a standard deviation of one, so that all coefficients can be read in units of one standard deviation. [2]

Figure 1 (Table 1) presents trends in average standardized math and reading scores by gender. Unlike the gap in reading scores that vanishes over the school years, the gap in math scores is already present at 7 years-old at the beginning of primary school (about 9% of a standard deviation), but after diminishing to approximately 4% of a standard deviation at 11 years of age, widens considerably at 16, at the end of secondary school, reaching almost 20 percent of a standard deviation.

[2] Figure A.1 in Appendix A shows the raw distributions of math and reading scores by age and gender. Even if all the children in our sample were born during the same week in March 1958, we include controls for when the tests were taken to allow for comparison of children of different ages.

2.2 Adolescence development

Our main control variables are measures of puberty development. The NCDS58 measures adolescence development at ages 11 and 16, through both medical assessments and direct questions to the child.

Our main measure of adolescent development is an index derived from the medical assessments carried out at ages 11 and 16. The *puberty index* compiles all the information in the medical assessments. Trained health personnel assessed anthropomorphic measures of the children at ages 11 and 16, using standard procedures as suggested by Brooks-Gunn and Warren (1987) and Coleman and Coleman (2002). At 11 years-old the NCDS58 information on the development of pubic hair for both girls and boys together with the development of breast for girls and genitalia for boys was gathered. All three development measures were coded from 1 (early development) to 5 (late development) by NCDS. As in Koerselman and Pekkarinen (2018), we add up the two corresponding measures for each child to obtain an increasing index of development ranging from 2 to 10 that we rescale to the 0-1 interval. Information on the development of pubic hair (1 absent; 2 sparse; 3 intermediate; 4 adult) and axillary hair (1 absent; 2 sparse; 3 intermediate; 4 adult) for girls and boys together with information on the development of facial hair for boys (1 absent; 2 sparse; 3 adult) and breast for girls (1 absent; 2 intermediate; 3 adult) was collected at age 16. Similarly, we add up the three corresponding measures for each child to obtain an increasing index of development ranging from 3 to 11 that we again rescale to the 0-1 interval. Therefore 0 corresponds to a child's adolescent development considered 'absent' in all measures; 1 corresponds to a child's development considered 'adult' in all measures; and the value of the index can be interpreted as the percentage of undergone pubertal maturation. Figure 2 presents the distribution of the indices by gender. We control for calendar month when assessments were performed to assure comparability of different individuals.

From children's questionnaires we know age at menarche for girls and when the voice broke for boys. Both the onset of the first menstruation for girls and voice breaking for boys constitute distinct events in puberty (Day et al. 2014). Age at menarche, even if not adequately validated, has been widely used in epidemiological studies as a proxy for pubertal stage (Baird et al 2017); conversely voice breaking is a validated marker of male puberty development, associated to Tanner stages 3 and 4 in late puberty (Hodges-Simeon et al 2013). In consequence our second measure of puberty development is *age at development*. Figure 2 also presents its distribution by gender.

Table 2 shows the descriptive statistics of our puberty measures by gender. Column 1 shows that at 11 years, both boys and girls are in the first stages of development, approximately in their first quarter. However, girls have an index of 0.2 points and boys a lower index of 0.15 points, in a 0-1 scale, that is, girls present on average a development level of 2 (early-to-mid-development) out of 5 possible stages for both breast development and pubic hair and boys may have a development level of 2 for genitalia development but just a development level of 1(early development) for public hair. At 16 years of age both boys and girls are in the second half of their puberty development (Column 2). On average, girls have completed 80 percent of their pubertal maturation (showing on average 'intermediate' hair and breast development), while boys have just completed about 60 percent (showing 'sparse' to 'intermediate' hair development). These disparities are also present when considering age at development, taking place on average for girls aged 12 and a half years and for boys aged almost 14 years (Column 3).

2.4 Controls and other variables used in the analysis

NCDS offers a vast array of potential control variables. We consider, not only demographic maternal characteristics (race, age at birth, lone parenthood, and any postsecondary education) usually included as controls in math gender gap analyses (see for instance Fryer and Levitt 2010), but also a whole bunch of additional variables that can safely be considered exogenous to the impact of puberty development on cognitive outcomes. These variables include several birth and pregnancy characteristics (birth order, birth weight, term pregnancy, birth complications, maternal smoking during pregnancy, mother working before birth), maternal and family background variables (English non-spoken at home, mother's height, mother's body mass index (BMI)), child's physical features (child's BMI at 7 years of age), schooling characteristics (class size, type of secondary school-grammar, modern, comprehensive), and region of residence at birth. Using such a rich set of control variables minimizes the risk of omitted variable bias when estimating skills' production functions. Descriptive statistics for these variables by gender are offered in Panels A through D of Table B.1 in Appendix B.

3. Empirical specification

Our goal is to determine whether there exists an effect of puberty development on boys' and girls' math performance. We model the math skills production function using the framework developed by Todd and Wolpin (2003, 2007), which considers school, family, and children's inputs.

Recent literature has emphasized the existence of sensitive periods in the formation of skills over the life cycle, that is, human capital inputs likely having different effects at different developmental stages (Cunha and Heckman 2007 Kautz et al. 2014, Keane et al. 2018). In particular, there is no a priori expectation that the impact of puberty development on math scores is constant with age. Therefore, as starting point, we test whether the effect of puberty development on math skills at 11 years of age is similar to the effect at 16 years of age, using the following contemporaneous model:

$$(1) \qquad Test_{it} = PU_{it}\alpha_t + B_i\rho_t + e_{it}$$

Where $Test_{it}$ is individual *i*'’s standardized test score at age t. PU_{it} is a vector of individual *i*'s degree of adolescence development at age t, as measured by one of our puberty measures; B_i is a vector of a wide set of controls, including pregnancy and birth characteristics, family features, school characteristics, and own child attributes as described in Table B.2 in Appendix B; and e_{it} is the error term. Notice that one of our puberty measures, age at development, is only available at age 16, so we only estimate equations for that age. Also, to allow for non-linearities we include a quadratic function of age at development in all the models.

As emphasized by Fiorini and Keane (2014), in estimating Equation (1) we face the problem of distinguishing a mere correlation between adolescence development and cognitive achievement from a true causal effect. A common source of endogeneity is simultaneity (reverse causality). This source of endogeneity is unlikely to appear in our case. Presumably more cognitively able children are not going to be able to influence their adolescence development. A second source of endogeneity relates to omitted variables, such as unobserved inputs and endowments. For instance, the child’s diet may influence both the child’s adolescence development and test scores. Given that no dataset contains a complete history of all relevant inputs, omitted variables remains problematic. The third source of potential endogeneity is measurement error in both inputs and outcome measures. We use a very rich source of data which hopefully reduces measurement error, but we also address this potential problem below.

The literature has proposed two broad estimation strategies to deal with endogeneity due to omitted variables: fixed-effect models and value-added models (see Todd and Wolpin 2003, 2007, Fiorini and Keane 2014, and Del Boca et al. 2017). Within child fixed effect (first difference) models require that human capital production functions are constant with age –so that both unobserved inputs and the endowment can be differenced out. This assumption is not plausible in the context of sensitive periods during childhood and adolescence. We thus turn to value-added models.

The value-added (VA) specification considers that both unobserved inputs and the endowment can be accounted for by past test scores. Todd and Wolpin (2003) show that this requires that the effects of all inputs, observed and unobserved, and the endowment decline at the same rate with age. In particular, for instance, the impact of puberty development at age 11 on test scores at age 16 would be a fraction λ_t of the impact of puberty development at age 11 on test scores at age 11 and the impact of the endowment at age 16 would be a fraction λ_t of the impact of the endowment at age 11. If the estimated coefficient of past test scores $\widehat{\lambda_t}$ is less than one, the value-added specification would imply that the effect of past inputs must always decrease over time. Formally:

(2) $$Test_{it} = PU_{it}\alpha_t + B_i\rho_t + \lambda_t Test_{it-1} + e_{it}$$

We can relax the assumption that the effect of observed inputs on test scores declines at the same rate that unobserved inputs and the endowment by estimating an extended version of the added value model that includes lagged inputs. We obtain the cumulative value-added (CVA) model (Todd and Wolpin 2007). Given that we only observe puberty development at 16 and 11 years, this model is only estimable for test scores at age 16. By including terms for past inputs in the estimating equation we are able to estimate long-term dynamic impacts of puberty development on cognitive skills and allow for different impacts of puberty development at age 11 on test scores at ages 16 and 11. We, however, still need to assume that the effect of unobserved inputs and endowments declines with age at a specific rate (λ_t) and is uncorrelated with the error term e_{it}. Formally:

(3) $$Test_{it} = \sum_{k=0}^{t} PU_{it-k}\alpha_k + B_i\rho_t + \lambda_t Test_{it-1} + e_{it}$$

Next, we address measurement error in the lagged test score.[3] In value-added models, measurement error tends to bias downwards the coefficient on lagged achievement (λ_t in Equation (3)) and may bias the observed input coefficients in an ambiguous direction (Del Bono et al. 2016,

[3] We do not address measurement error in puberty development. First, because doctors' assessments are less likely to show measurement errors correlated to test scores; and second, because it is difficult to obtain valid instruments for puberty measures.

Keane et al. 2018). Our preferred specification involves instrumenting for lagged test scores using two-period lagged outcomes in a cumulative value-added instrumental variables (CVA-IV) specification:

$$(4) \qquad Test_{it} = \sum_{k=0}^{t} PU_{it-k}\alpha_k + \sum_{k=0}^{t} X_{it-k}\beta_k + B_i\rho_t + \lambda_t \widehat{Test_{it-1}} + e_{it}$$

Notice that this model can only be estimated for skills at age 16 and that when using the quadratic function of age at development it tuns into a value-added instrumental variables model.

4. The causal dynamic effects of puberty development and Math test scores

In this section we present our main estimation results for the impact of puberty development on math performance of boys and girls. We evaluate the impact of puberty development at 11 and 16 years old on math scores at 11 and 16 years of age, using the contemporaneous, VA, CVA, and CVA-IV models of Section 3.

The different panels in Table 3 estimate the relationship between puberty and Math test scores by child's age and gender using the contemporaneous specification in Equation (1) for the medical measures of puberty development: the puberty indices at ages 11 and 16. It clearly shows different associations between puberty development and math test scores, by gender and age: whereas boys show similar positive associations at 11 and 16 years of age, for girls the positive associations at 11 years of age are no longer present at 16 years. For instance, Column 1 shows that a one-point increase in the puberty index at age 11, that is, going from totally immature to completely developed, is associated with increases in boys' Math scores of 20% of a standard deviation. For a more realistic change in boys' puberty index of 10 percentage points at age 11 (such as presenting genitalia development at stage 2 instead of 1 out of 5 possible stages) Math test scores increase by approximately 2% of a standard deviation at 11 years of age, an impact similar to 6% of the impact of having a mother with post-secondary education. Column 2 shows that a similar 10-percentage-point increase in the puberty index at age 11 for girls (such as presenting breast development at stage 2 instead of 1 out of 5 possible stages) is associated with increases in Math test scores at age 11 of approximately 1.7% of a standard deviation, an impact similar to 5% of the impact of having a mother with post-secondary education. Similarly, Columns 3 and 4 show that a 10 percentage point increase in the puberty index at 16 years of age (such as presenting adult pubic hair development at stage 4 instead of stage 2 –sparse- out of 4 possible stages) is associated with increases in boys' Math tests scores at age 16 of approximately 2% of a standard deviation again, but shows no correlation with girls' scores at that age.

The results in Columns 1 to 4 of Table 4 address potential endogeneity concerns of omitted variables and measurement error by estimating the value-added models of Equations (2) to (4). Overall, controlling for previous performance reduces the contemporaneous impact of puberty development on boys' and girls' math test scores and even turns the impact negative for 16-year-old girls. The fact that the estimates of the value-added model in Table 4 differ from the estimates of the contemporaneous model in Table 3 suggest that unobserved inputs and the endowment are correlated with puberty development and the estimates in Table 3 are likely biased.

Columns 1 and 2 of Table 4 show that at 11 years of age, the impact of current puberty development is no longer significant for boys and drops to half of its size for girls, compared to the figures in Table 3. That is, a 10pp increase in girls' puberty index at 11 increases math scores just 0.9% of a standard deviation. Also, at 16 years of age (columns 3 and 4), a 10pp increase in the puberty index at age 16 (such as presenting adult axillary hair development at stage 4 instead of stage 3 –intermediate- out of 4 possible stages) increases boys' math scores by about 0.8% of a standard deviation, but reduces girls' math scores by 0.9% of a standard deviation. Therefore, moving from absent (0) to mature (1) puberty development at age 16 increases boys' math scores and decreases girls' math test scores by roughly 10% of a standard deviation.

Columns 5 and 6 of Table 4 present estimation results for the cumulative value-added model of Equation (3) that relax the assumption that the effect of puberty development on math scores declines over time at the same rate as the impacts of unobserved inputs and the endowment. In general, including information on puberty development at age 11 does not change the contemporaneous impact of puberty development on boy's math scores at age 16, but further decreases the contemporaneous impact of puberty on girls' math scores at age 16. Given that the value-added model is nested on the cumulative value-added model, these results imply that the value-added model may be adequate to estimate the unbiased impact of puberty development on boys' math skills, but not on girls' math skills. In particular, columns 5 and 6 of Table 4 show that at 16 years of age, a 10pp increase in the puberty index at age 16 increases boys' math scores by about 0.8% of a standard deviation as in Column 3, but reduces girls' math scores by 1.2% of a standard deviation, 0.3pp more than previously in Column 4. Column 6 additionally shows that puberty development has long-term effects beyond its impact through past scores for girls. A 10-percentage-point increase in the puberty index of girls at age 11 (such as presenting breast development at stage 3 instead of 2 out of 5 possible stages) increases Math test scores at age 16 by 1.4% of a standard deviation, that is, roughly a 15% increase for moving from infant (0) to mature (1) development at age 11.

Columns 7 and 8 of Table 4 further address measurement problems. Using twice-lagged skills to instrument for lagged test scores increases the coefficient of the lagged tests scores as expected by about 20 to 30 percent. The point estimates of the effect of puberty development on math scores remain very similar, however. For boys, a 10pp increase in the puberty index at age 16 increases boys' math scores by about 0.6% of a standard deviation, significant at the 17% level, slightly lower than reported results in Columns 3 and 5. A similar increase in the puberty index at age 16 reduces girls' math scores by 1.4% of a standard deviation, 0.2pp more than previously in Column 6. The long-term impact of a 10-percentage-point increase in the puberty index of girls at age 11 increases Math test scores at age 16 by 1.3% of a standard deviation, 0.1pp less than in Column 6.

To put these results into perspective, Carneiro et al (2013) report for instance that one additional year of mother's education increases mathematics standardized scores by 3.8% of a standard deviation at 12 to 14 ages. Our estimates imply that the impact of puberty development on math scores at 16 years of age, that is, the impact of moving from infant to mature in the puberty index at 11 years for girls (about 15% of a standard deviation), or moving from infant to mature in the puberty index 16 years for boys (roughly 7% of a standard deviation), may be similar to the impact of between 1 and 4 more years of maternal education. The impact of puberty development on math scores is quantitatively and qualitatively important.

In other to see the contribution of puberty to explaining the math gender gap we estimate equations 2, 3, and 4 by age including a female dummy indicator and its interaction with our puberty measure. Estimating one equation for boys and girls implies assuming that boys and girls share the same production function (see the discussion in Baker and Milligan 2016). Table 5 presents the results for the change in the math gender gap when we control for puberty development. To get a sense of the impact of controlling for puberty development, we first show the size of the gender gap for the different models in the previous Table 4 before any puberty measure or its interaction with gender are included (Panel A). The specifications in Panel B of Table 5 include the puberty index and its interaction with the female dummy. Once previous scores are controlled for, there is virtually no gender gap in maths at 11 years of age. Including controls for puberty development does not change this finding (Column 1 in Table 5). However, at 16 years of age there is a gap of about 18% of a standard deviation (Columns 2 to 4 in Panel A) that diminishes to almost a third of its previous size and turns insignificant when puberty development and its interaction are included in the regressions (Columns 2 to 4 in Panel B).

4.1. Robustness Checks to Alternative Model Specifications and Measures of Adolescence Development

In this section we assess the sensitivity of our findings to (1) changes in model specification and (2) using an alternative self-assessed measure of puberty development. We first present two departures from our basic value-added model for the impact of puberty development at 16 years of age that consider additional information on reading scores and non-cognitive skills in estimating math skills production functions. First, Del Boca et al (2017) propose an additional model that relaxes the requirement that past scores are uncorrelated with the error term in Equation (3). In their two-step estimation procedure, they first compute an individual fixed effects estimation of the depreciation rate (λ_t in Equation (3)), using the information on the different skills for each child at times t and t-1. We use math, reading, and non-cognitive test scores at 16 and 11 years of age to estimate this persistence parameter that controls for the child-specific endowment. In the second step, λ_t is replaced with its estimate from the fixed effects regression, obtaining the two-step cumulative value-added model:

$$(5) \qquad Test_{it} - \widehat{\lambda_t} Test_{it-1} = \sum_{k=0}^{t} PU_{it-k}\alpha_k + \sum_{k=0}^{t} X_{it-k}\beta_k + B_i\rho_t + e_{it}$$

Secondly, in the spirit of Cunha et al (2010) we consider skills as latent variables imperfectly measured and include past measures of other skills, reading and non-cognitive skills, in the production function of math skills. Following Agostinelli et al (2019) we further address potential mismeasurements in all lagged skills measures by instrumenting them with two-period-lagged skills measures. Our second robustness check involves estimating a variant of Equation (4) that incorporates information on past reading and non-cognitive outcomes, adequately instrumented:

$$(6) \qquad Test_{it}^{k} = \sum_{k=0}^{t} PU_{it-k}\alpha_k + \sum_{k=0}^{t} X_{it-k}\beta_k + B_i\rho_t + \lambda_t \sum_{j=1}^{J} \widehat{Test_{it}^{J}} + e_{it}$$

Panel A in Table 6 shows that the conclusions from the CVA-IV model of Columns 7 and 8 of Table 4 (reproduced here in Columns 1 and 2 in Panel A of Table 6) remain when estimating both the two-step and the latent-factors CVA models of Equations (5) and (6). In particular, the contemporaneous impact of puberty development on boys' math scores at 16 years of age ranges now from 0.6% to 1.2% of a standard deviation for a 10pp increase in the puberty index. Similarly, a 10pp increase in girls' puberty index at 16 years of age decreases girls' math scores between 1 and 2% of a standard deviation, and a similar increase in girls' puberty index at 11 years of age increases girls' math scores by between 1.3 and 2% of a standard deviation.

Subsequently, we test the sensitivity of our findings to using self-reported age at development as an alternative measure of adolescent development. As previously stated, while voice breaking has been validated as an adequate marker for male puberty development, girls age at menarche has not. Panel B of Table 5 looks at the association of the quadratic function of age at development and Math scores at age 16 using the CVA-IV model of Equation (4). Age at development has a convex – though non-significantly different from zero– relationship with girls' math scores and a clearly concave relationship with boys' math scores. This relationship with boys' math scores presents a maximum for between 13.1 and 13.4 years of age, about half a year before the average age at development of boys of 13.9 years. The estimated impact implies that for the representative boy developing at 13.9 years of age, developing 6 months earlier accounts for between 0.3% and 0.7% of a standard deviation higher math score.

Taken together, the results in this section suggest that our findings from Table 4 are quite robust and the assumptions made in Equation (4) can be considered a lower bound for the impact of puberty development on math scores by boys and girls.

4.2. Identification Checks: Looking for Any Remaining Sources of Bias

The main assumption behind the value-added models used in the previous analysis is that, after controlling for past test scores, there remain no unobserved factors affecting the impact of puberty development on math performance. There is no formal way of testing this assumption. In what follows we present suggestive evidence that any unobserved factors, such as mothers' IQ or nutrition patterns, must be uncorrelated with puberty development once past scores and the vast set of covariates are controlled for. The exercise consists of identifying a set of variables, birth weight, maternal education, and smoking during pregnancy, that could be related to these unobserved factors. For instance, birth weight has been shown to have long-lasting impacts on cognitive development (Figlio et al 2014). Similarly, maternal education also increases the child's cognitive ability (Carneiro et al 2013) and maternal smoking during pregnancy is associated with lower cognitive test scores and behavioural problems (Wehby et al 2011). If these variables are measured before the adolescence period, they cannot possibly be affected by puberty development unless there remains some selection bias not properly accounted for by model specifications.

In Table 7 we estimate the models in Equations 2 to 4 using birth weight, maternal education, and smoking during pregnancy as dependent variables. The results therein basically show that puberty development is virtually unrelated to pre-determined variables.

We also control for a variety of covariates that have been shown to be correlated with puberty as well as math outcomes and show that neither family background, nor school characteristics, nor the physical and psychological features of the child affect the estimated impact of puberty development on boys' and girls' math scores.[4] Family background may impact gender differences in math skills through at least three potential channels. First, family resources may impact differently boys' and girls' educational outcomes (Autor et al. 2019, Lundberg 2017). Second, parents may make differentiated gender-specific investments, particularly time investments (Baker and Milligan 2016). For instance, sons may receive more time due to the extra time devoted by their fathers (Lundberg 2005). And third, parental preferences or gender attitudes may be transmitted to their children (Dossi et al 2019). There is evidence on the critical impact of same-sex family figures during adolescence in shaping later roles and choices (Olivetti et al 2018). The literature shows not only correlations between mothers' and children's attitudes towards working women (Farre and Vella 2013), but also correlations between mothers' and daughters' labour supply (Olivetti et al. 2018). Fryer and Levitt (2010) hypothesize that maternal occupation in STEM jobs and expectations for having children in STEM jobs should reduce the gender gap in math but find no corroborating evidence. Table B.4 shows that family background does not affect the impact of puberty development on boys' and girls' math achievement. Controlling for maternal labour supply, occupation, and expectations do not significantly alter the impact of puberty development on math scores.

School environments may also affect the gender gap in math, especially through the influence of, first, teachers, and, second, peers. Teachers gender role beliefs and expectations may influence

[4] Table B.3 In Appendix B offers descriptive statistics by gender for the variables used in this section.

students' self-image, affecting their interests and aspirations and ultimately their academic outcomes, leading to a self-fulfilling prophecy (Papageorge et al 2019, Carlana 2019). Also, teachers may interact differently with boys and girls - giving different types of feedback or spending more time training boys in math relative to girls - directly affecting their academic outcomes (Lavy and Sand, 2018, Alan et al. 2018). Additionally, the gender composition of the classroom has often been studied as a potential explanatory factor in gender differences in math scores. Females are more likely to conform with gender stereotypes in the presence of males to avoid disappointing gender-specific expectations (Steele 1997). Alternatively, girls may also shy away from competition, especially when confronted with boys (Gneezy et al. 2003, Niederle and Vesterlund 2007, 2010). The empirical evidence shows that the proportion of boys in the classroom does not affect the gender gap in math, however, except for when it is restricted to the proportion of high achieving boys (Bharadwaj et al. 2015, Cools et al. 2019). Nonetheless, single-sex schooling has been found to improve girls' academic outcomes in math and math related subjects (Eisenkopf et al. 2015, Booth et al. 2018). Table B.5 shows that school inputs do not significantly change the estimated impact of puberty on math scores of boys and girls. Neither teacher gender, nor teacher expectations, nor the number of math courses taken, nor co-educational schooling change the estimated impact of puberty development on math scores.

Own-child factors may also influence the gender gap in math. First, from a biological standpoint overweight has been shown to be correlated with academic outcomes (Sabia 2007). Second, differences in perseverance, risk aversion, or leadership may also be behind math differences between boys and girls. For instance, Elison et al. (2018) show that girls give up more easily; Andreoni et al. (2019) show that they are more risk averse; and Alan et al (2019) show that they are less willing to take leadership roles. Table B.6 shows that neither the physical nor the psychological features of the child alter the estimated impact of puberty development on math scores. Both overweight children and children with behavioural problems seem to show similar impacts of puberty measures on math achievement.

5. Understanding the mechanisms behind the impact of puberty on math outcomes by gender

We have seen that puberty changes may explain if not completely, at least partially the widening of the gender gap in math. In this section we shed some light on the mechanisms that may be underlying this relationship. In what follows we present several pieces of evidence that support that culture and socialization appear to interact with biological forces in shaping how adolescent development impacts the math performance of boys and girls.

We first draw the attention towards the different impacts of puberty development by age. If puberty development impacted math outcomes purely through a physical mechanism its impact should not change with the child's age. The fact that we find different impacts of puberty development at 11 and 16 years, especially for girls, in Tables 3 and 4 suggests a social channel coming into play alongside the physical changes.

Second, heterogeneity analyses by initial skills, and self-assessed math performance also point at the direction of self-confidence affecting how puberty development impacts math performance. There is some consensus in the literature that the gender gap in math tends to be larger at the upper tail of the test scores distribution, as boy's variability in results is higher than girls' (Machin and Pekkarinen 2008, Hyde et al. 2008, Ellison and Swanson 2010).[5] For instance, using data from the American Mathematical Competitions, Ellison and Swanson (2010) document a 4.1 to 1 male to female ratio for scoring at the 33% top of the distribution and show that this gender gap at high achievement levels increases for higher percentiles in the distribution. Also, differences in self-confidence in own abilities in math-related tasks may also impact the gender gap in math. For instance, Coffman (2014) shows that women report significantly less confidence in gender-incongruent than gender-congruent tasks. Bordalo et al. (2019) further distinguish the role of social stereotypes from the role of self-confidence showing a significant impact of self-confidence on the math gender gap. There is also some evidence pointing at an increase in self-esteem upon development for boys, and a decrease for girls (Martin and Steinbeck 2017).

To explore potential heterogeneity in the impact of puberty development by initial skills, and self-assessed math performance, we re-estimate the VA specifications and include interactions to determine if the effect of puberty development is different for girls and boys with higher math skills at 7 years of age, and above average level of self-assessed math performance. The estimates in Panel A of Table 8 show that there are no significant differences in the impact of puberty development on math scores by initial skills. Panel B, however, reveals that the positive impact of puberty

[5]. There is some controversy regarding this greater variability hypothesis, though. Using PISA data for 2003, Machin and Pekkarinen (2008) find that the average variance ratio (boys' variance over girls' variance) which should be 1.00 if there was equality is 1.13; and is even larger -1.70- when considering the top 5 percent of the distribution. However, Using TIMMs data for 2003 and 2007, Kane and Mertz (2012) show evidence against this greater male variability hypothesis: the average variance ratio ranges from 0.90 to 1.49 across countries and therefore the greater variability of boys' results over girls' is not found in some countries.

development on boys' math scores is driven by those with higher self-assessed math efficacy. In particular, for boys with above average self-efficacy a 10pp. increase in the puberty index at age 16 is associated with almost a 2% of a standard deviation increase in math scores. Even if there are no differences for girls of high and low self-perceived math ability, puberty development impacts positively math scores of high self-perceived ability boys, though not those of low self-perceived ability boys, These observed differences by self-perceived ability seem to rule out that the effect from puberty on math outcomes by gender purely come from biological changes.

Third, social stereotypes dictate that men are good at math and women are good at verbal skills (Bordalo et al 2016, 2019). If puberty development impacted math outcomes through a social mechanism, we should expect to see mirror images of math performance on English performance by boys and girls: that is, puberty development should increase girls' English performance and boys' math performance while harming girls' math performance and boys' English performance. Table 9 shows that puberty development benefits English outcomes of both boys and girls. Therefore, social stereotypes might be behind the changes in the math performance of girls, but not boys.

All in all, the evidence gathered regarding differentiated impacts of puberty development on achievement by age and subject for girls and by self-perceived ability for boys seem to rule out a purely biological explanation for the gender gap in maths and suggests that social conditions interact with biological factors in shaping gender gaps in maths.

6. Conclusion

This paper investigates the causal relationship between adolescence development and the increase in the gender gap in math performance from primary to secondary school. We use a unique dataset that followed all British children born in the first week of March 1958 (the 1958 NCDS) and offers medical assessments of puberty development alongside cognitive outcomes and a rich set of household, school, and birth characteristics. We document that the gender gap in math widens by about 10% of a standard deviation during the adolescence years. Using dynamic production function models, we show that adolescent development can explain almost two thirds of the increased gap in math performance between boys and girls. Our results are robust to using different alternative model specifications and a different self-assessed measure of adolescent development.

We also explore the mechanisms behind the estimated impact of adolescent development on the gender gap in math performance. The fact that impacts vary depending on the age of boys and girls, the specific subject under study: math or reading, and the degree of self-perceived math ability allow us to rule out a purely biological channel for this effect. Consistent with recent neurobiology theories that emphasize that brain synapses are pruned in response to environmental influences (Dahl et al. 2019), our results imply a combined socio-biological explanation for the gender gap in maths. More research is clearly needed in this area.

Acknowledgements

Cristina Borra acknowledges support by the Spanish Ministry of Science, Innovation, and Universities (RTI2018-098217-B-I00). Almudena Sevilla acknowledges support by the European Research Council (ERC-2017-COG, Grant agreement ID: 770839).

References

Agostinelli, Francesco & Morteza Saharkhiz & Matthew Wiswall, 2019. "Home and School in the Development of Children," Working Papers 2019-044, Human Capital and Economic Opportunity Working Group.

Akerlof, G.A. and R.E. Kranton. 2000. "Economics and identity." Quarterly journal of Economics 115 (3):715–753.

Alan, Sule & Seda Ertac & Ipek Mumcu, 2018. "Gender Stereotypes in the Classroom and Effects on Achievement," The Review of Economics and Statistics, vol. 100(5), pages 876-890.

Alan, Sule, Seda Ertac, Elif Kubilay, Gyongyi Loranth, Understanding Gender Differences in Leadership, The Economic Journal, Volume 130, Issue 626, February 2020, Pages 263–289

Andersen, Steffen & Seda Ertac & Uri Gneezy & John A. List & Sandra Maximiano, 2013. "Gender, Competitiveness, and Socialization at a Young Age: Evidence From a Matrilineal and a Patriarchal Society," The Review of Economics and Statistics, vol. 95(4), pages 1438-1443.

Andreoni, J., Di Girolamo, A., List, J.A., Mackevicius, C., and Samek, A. (2019) Risk preferences of Children and Adolescents in Relation to Gender, Cognitive Skills, Soft Skills, and Executive Functions. NBER Working Paper No. 25723.

Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman (2019). "Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes." American Economic Journal: Applied Economics 11(3): 338–381.

Baird J, Walker I, Smith C, Inskip H. Review of methods for determining pubertal status and age of onset of puberty in cohort and longitudinal studies. London, UK: CLOSER; 2017. Available from: https://www.closer.ac.uk/wp-content/uploads/Review-of-methods-for-determining-pubetal-status-FULL-FINAL.pdf

Baker, Michael, and Kevin Milligan. 2016. "Boy–Girl Differences in Parental Time Investments: Evidence from Three Countries." Journal of Human Capital 10 (4): 399–441

Bharadwaj, Prashant & Giacomo De Giorgi & David Hansen & Christopher A. Neilson, 2016. "The Gender Gap in Mathematics: Evidence from Chile," Economic Development and Cultural Change, vol. 65(1), pages 141-166.

Black, Dan A., Amelia M. Haviland, Seth G. Sanders, and Lowell J. Taylor. 2008. "Gender Wage Disparities among the Highly Educated." Journal of Human Resources 43 (3): 630–59.

Blakemore SJ, Choudhury S. Development of the adolescent brain: implications for executive function and social cognition. J Child Psychol Psychiatry. 2006 Mar-Apr;47(3-4):296-312.

Blakemore, S.-J. (2018). Inventing ourselves: The secret life of the teenage brain. New York, NY: PublicAffairs.

Blau, Francine D., and Lawrence M. Kahn. 2017. "The Gender Wage Gap: Extent, Trends, and Explanations." Journal of Economic Literature 2017, 55(3), 789–865.

Booth, Alison L. & Cardona-Sosa, Lina & Nolen, Patrick, 2018. "Do single-sex classes affect academic achievement? An experiment in a coeducational university," Journal of Public Economics, Elsevier, vol. 168(C), pages 109-126.

Bordalo Pedro & Katherine Coffman & Nicola Gennaioli & Andrei Shleifer, 2019. "Beliefs about Gender," American Economic Review, 109(3), 739-773.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. "Stereotypes." Quarterly Journal of Economics 131, no. 4 (November 2016): 1753–1794.

Borgonovi, F., A. Choi, and M. Paccagnella. 2018. The Evolution of Gender Gaps in Numeracy and Literacy between Childhood and Adulthood. Paris: OECD Education Working Papers

Brooks-Gunn J, Warren MP, Rosso J, Gargiulo J. (1987). Validity of self-report measures of girls' pubertal status. Child Development 58(3):829-41.

Card, David, and A. Abigail Payne, 2017. "High School Choices and the Gender Gap in STEM," NBER Working Paper, 2017.

Carlana, Michela 2019 Implicit Stereotypes: Evidence from Teachers' Gender Bias, The Quarterly Journal of Economics, 134, 1163–1224.

Carneiro, Pedro, Costas Meghir, Matthias Parey. 2013. Maternal Education, Home Environments, and the Development of Children and Adolescents, Journal of the European Economic Association, Volume 11, Issue suppl_1, Pages 123–160.

Cavanagh, S., Riegle-Crumb, C., Crosnoe R., 2007, "Puberty and the Education of Girls." Social Psychology Quarterly, 70:2, 186-198.

Coffman, Katherine Baldiga 2014. Evidence on Self-Stereotyping and the Contribution of Ideas , The Quarterly Journal of Economics, Volume 129, Issue 4, Pages 1625–1660.

Coleman L, Coleman J. The measurement of puberty: a review. J Adolesc. 2002 Oct;25(5):535-50.

Contini, Dalit & Tommaso, Maria Laura Di & Mendolia, Silvia, 2017. "The gender gap in mathematics achievement: Evidence from Italian data," Economics of Education Review, vol. 58(C), pages 32-42.

Cools, Angela, Raquel Fernández, and Eleonora Patacchini, "Girls, Boys, and High Achievers," NBER Working Paper 25763, 2019.

Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." American Economic Review, 97 (2): 31-47.

Dahl, R., Allen, N., Wilbrecht, L. et al. (2018). Importance of investing in adolescence from a developmental science perspective. Nature 554, 441–450.

Day, F., Elks, C., Murray, A. et al. 2015. Puberty timing associated with diabetes, cardiovascular disease and also diverse health outcomes in men and women: the UK Biobank study. Sci Rep 5, 11208.

Del Boca, Daniela, Chiara Monfardini, and Cheti Nicoletti. 2017. Parental and Child Time Investments and the Cognitive Development of Adolescents. Journal of Labor Economics 2017 35:2, 565-608

Del Bono, Emilia & Marco Francesconi & Yvonne Kelly & Amanda Sacker, 2016. "Early Maternal Time Investment and Early Child Outcomes," Economic Journal, vol. 126(596), pages 96-135.

Dossi, G., Figlio, D. N., Giuliano, P., & Sapienza, P. (2019). Born in the Family: preferences Preferences for boys and the gender gap in math. NBER Working Paper No. 25535.

Dreber, Anna & von Essen, Emma & Ranehill, Eva, 2011. "Age at pubertal onset and educational outcomes," Research Papers in Economics 2011:26, Stockholm University, Department of Economics.

Eisenkopf, Gerald & Hessami, Zohal & Fischbacher, Urs & Ursprung, Heinrich W., 2015. "Academic performance and single-sex schooling: Evidence from a natural experiment in Switzerland," Journal of Economic Behavior & Organization, vol. 115(C), pages 123-143.

Ellison, G., and A. Swanson. 2010. "The Gender Gap in Secondary School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions." Journal of Economic Perspectives 24 (2):109–28.

Ellison, Glenn and Ashley Swanson, 2018 "Dynamics of the Gender Gap in High Math Achievement," NBER Working Paper 24910.

Figlio, David, Jonathan Guryan, Krzysztof Karbownik, and Jeffrey Roth. 2014. "The Effects of Poor Neonatal Health on Children's Cognitive Development." American Economic Review, 104 (12): 3921-55.

Fiorini, Mario, and Michael P. Keane. 2014. "How the allocation of children's time affects cognitive and noncognitive development." Journal of Labor Economics, 32, no. 4:787-836.

Fryer, R., and Levitt, S. (2010). An empirical analysis of the gender gap in mathematics. American Economic Journal: Applied Economics, 2(2), 210– 240.

Fuhrmann D, Knoll LJ, Blakemore SJ. Adolescence as a Sensitive Period of Brain Development. Trends Cogn Sci. 2015 Oct;19(10):558-566.

Giedd, J.N., Blumenthal, J., Jeffries, N.O., Castellanos, F.X., Liu, H., Zijdenbos, A., Paus, T., Evans, A.C., Rapoport, J.L., 1999. Brain development during childhood and adolescence: a longitudinal MRI study. Nat. Neurosci. 2, 861–863.

Gneezy, U., Niederle, M. and Rustichini, A. (2003). Performance in competitive environments: Gender differences. The Quarterly Journal of Economics, 118 (3), 1049–1074.

Goddings, A. L., Mills, K. L., Clasen, L. S., Giedd, J. N., Viner, R. M., & Blakemore, S. J. (2014). The influence of puberty on subcortical brain development. NeuroImage, 88, 242– 251.

Goldin, C., Katz, L. F., and Kuziemko, I. (2006). The homecoming of American college women: The reversal of the college gender gap. Journal of Economic Perspectives, 20(4):133{156.

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, Gender, and Math. Science 320, 1164–1165.

Guiso, Luigi, Ferdinando Monte, Paola Sapienza, and Luigi Zingales. 2008. “Culture, Gender, and Math.” Science, 320(5880): 1164–65.

Haynie, D. L. (2003). Contexts of risk? Explaining the link between girls’ pubertal development and their delinquency involvement. Social Forces, 82(1), 355–397.

Herting MM, Gautam P, Spielberg JM, Dahl RE, Sowell ER. 2015. A longitudinal study: changes in cortical thickness and surface area during pubertal maturation. PLoS One 10(3)

Hodges-Simeon CR, Gurven M, Cárdenas RA, Gaulin SJ. 2013. Voice change as a new measure of male pubertal timing: a study among Bolivian adolescents. Ann Hum Biol. 40(3):209-19

Hyde, J. S., and Mertz, J. E. (2009). Gender, culture, and mathematics performance. Proceedings of the National Academy of Sciences of the United States of America, 106(22), 8801–8807.

Joensen, J. S. and Nielsen, H. S. (2016), “Mathematics and Gender: Heterogeneity in Causes and Consequences”, The Economic Journal, 126/593, 1129-1163.

Kahn, S. and D. Ginther (2017). “Women and STEM”. National Bureau of Economic Research Working Paper.

Kane, Jonathan M. and Janet E. Mertz, (2012) “Debunking Myths about Gender and Mathematics Performance.” Notices of the AMS, January.

Kautz, T., et al. (2014), "Fostering and Measuring Skills: Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success", OECD Education Working Papers, No. 110.

Keane, Michael P & Sonya Krutikova & Timothy Neal, 2018. "The impact of child work on cognitive development: results from four Low to Middle Income countries," IFS Working Papers W18/29.

Koerselman, Kristian & Pekkarinen, Tuomas, 2018. "Cognitive consequences of the timing of puberty," Labour Economics, vol. 54(C), pages 1-13

Koivusilta, L., & Rimpelä, A. (2004). Pubertal timing and educational careers: a longitudinal study. Annals of Human Biology, 31(4), 446–465.

Lavy, Victor, and Edith Sand, (2018) “On the Origins of Gender Human Capital Gaps: Short and Long Term Consequences of Teachers’ Stereotypical Biases,” Journal of Public Economics, 167, 263–279.

Lídia Farré & Francis Vella, 2013. "The Intergenerational Transmission of Gender Role Attitudes and its Implications for Female Labour Force Participation," Economica, vol. 80(318), pages 219-247.

Lundberg, Shelly, 2017. "Father Absence and the Educational Gender Gap," IZA Discussion Papers 10814

Machin, S. & Pekkarinen, T. (2008). Global sex differences in test score variability. Science, 322, 1332–1333.

Martin, A. J., & Steinbeck, K. (2017). The role of puberty in students' academic motivation and achievement. Learning and Individual Differences, 53, 37–46.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). TIMSS 2015 International Results in Mathematics. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2015/international-results/

Niederle, M., and Vesterlund, L. (2010). Explaining the Gender Gap in Math Test Scores: The Role of Competition. Journal of Economic Perspectives, 24(2), 129–144.

Nollenberger, Natalia, Núria Rodríguez-Planas, and Almudena Sevilla. 2016. "The Math Gender Gap: The Role of Culture." American Economic Review, 106 (5): 257-61.

OECD (2010), PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Volume I), PISA, OECD Publishing, Paris, https://doi.org/10.1787/9789264091450-en.

OECD (2014), PISA 2012 Results: What Students Know and Can Do (Volume I, Revised edition, February 2014): Student Performance in Mathematics, Reading and Science, PISA, OECD Publishing, Paris, https://doi.org/10.1787/9789264208780-en.

OECD (2018), Education at a Glance 2018: OECD Indicators, OECD Publishing, Paris, https://doi.org/10.1787/eag-2018-en.

OECD (2019a), Adult education level (indicator). doi: 10.1787/36bce3fe-en (Accessed on 10 October 2019)

OECD (2019b), PISA 2018 Results (Volume II): Where All Students Can Succeed, PISA, OECD Publishing, Paris, https://doi.org/10.1787/b5fd1b8f-en.

Olivetti, Claudia, Eleonora Patacchini, Yves Zenou, (2018). "Mothers, Peers, and Gender-Role Identity", Journal of the European Economic Association, jvy050, https://doi.org/10.1093/jeea/jvy050.

Papageorge, NicholasW., Seth Gershenson, and Kyungmin Kang, 2019 "Teacher Expectations Matter," Review of Economics and Statistics, https://doi.org/10.1162/rest_a_00838.

Picci G, Scherf KS. 2016. From Caregivers to Peers: Puberty Shapes Human Face Perception. Psychol Sci. 27(11):1461-1473.

Piekarski, D. J. et al. 2017 Does puberty mark a transition in sensitive periods for plasticity in the associative neocortex? Brain Res. 1654, 123–144.

Rose, H. and Betts, J. (2004). 'The effect of high school courses on earnings', Review of Economics and Statistics, vol. 86(2), pp. 497–513.

Steinberg L, Graham S, O'Brien L, Woolard J, Cauffman E, Banich M. 2009. Age differences in future orientation and delay discounting. Child Dev. 80(1):28-44.

Steinberg, L. (2014) Age of Opportunity – Lessons from the New Science of Adolescence, New York: Houghton Mifflin Harcourt.

Todd, Petra, and Ken Wolpin. 2003. On the specification and estimation of the production function for cognitive achievement. Economic Journal, 113:F3-F33.

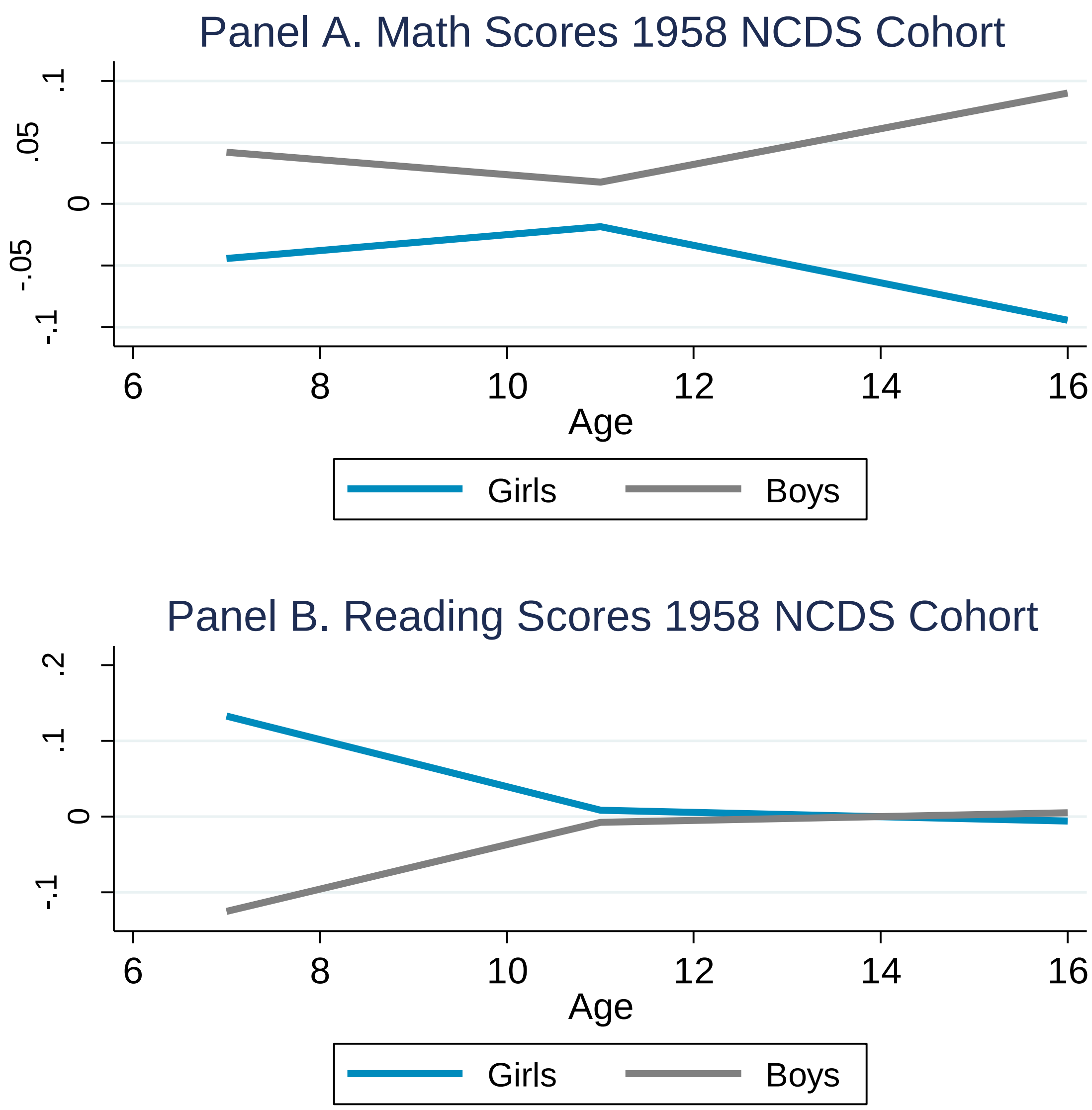
Todd, Petra, and Ken Wolpin. 2007. The production of cognitive achievement in children: home, school and racial test score gaps. Journal of Human Capital, 1:91-136.

Wehby, G.L., Kaitlin, P., McCarthy, A.M., Castilla, E., Murray, J.C., 2011. The impact of maternal smoking during pregnancy on early child neurodevelopment. Journal of Human Capital 5, 207–254.

Figures

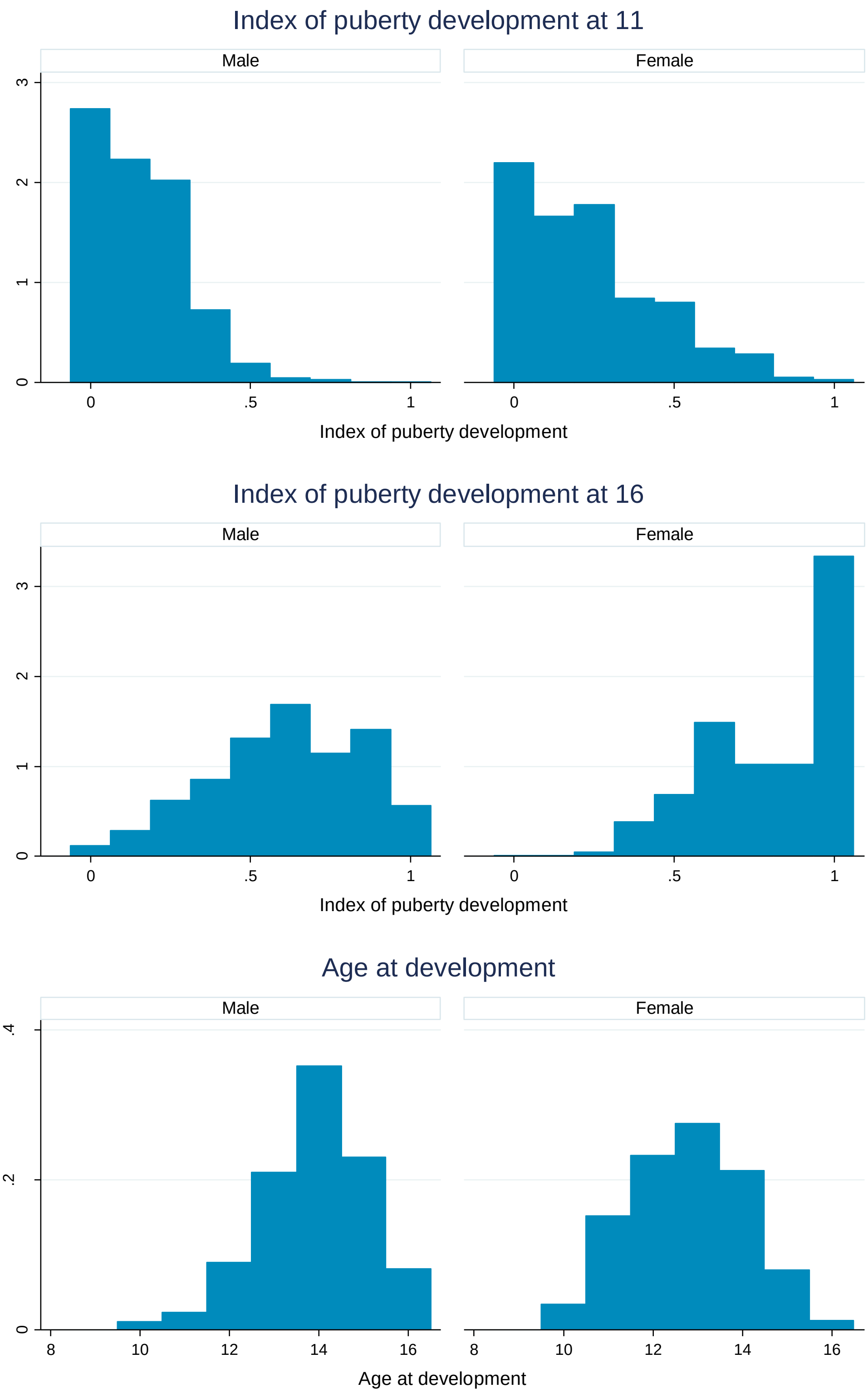
Figure 1. Standardized Mathematical Scores by Gender over the lifetime.

1958 NCDS cohort.



Source: NCDS Longitudinal Data.

Figure 2. Measures of Puberty Development by Gender over the Lifetime. 1958 NCDS cohort.



Source: NCDS Longitudinal Data

Tables

Table 1. Standardized Math and Reading Scores. 1958 NCDS cohort.

	At 7 years	At 11 years	At 16 years
Panel A. Math Scores			
Female	-0.0443	-0.0186	-0.0946
	(0.994)	(0.974)	(0.949)
Male	0.0420	0.0176	0.0901
	(1.004)	(1.024)	(1.038)
Mean difference significant	-0.086***	-0.036*	-0.185***
Nobs.	14897	14126	11920
Panel B. Reading Scores			
Female	0.1329	0.0083	-0.0058
	(0.938)	(0.955)	(0.963)
Male	-0.1256	-0.0078	0.0055
	(1.040)	(1.040)	(1.035)
Mean difference significant	0.259***	0.016	-0.011
Nobs.	14929	14130	11986

Notes: The table shows means by gender. Standard deviations in parenthesis. Mean differences significant at the * 10%, ** 5 % and *** 1% levels.

Source: NCDS Longitudinal Data.

Table 2. Descriptive statistics: Puberty measures

	(1)	(2)	(3)
	Puberty Index		Age at development
	At 11 years	At 16 years	
Female	0.2345	0.8036	12.7705
	(0.218)	(0.206)	(1.320)
Male	0.1518	0.6096	13.8889
	(0.146)	(0.241)	(1.226)
Mean diff. significant	-0.083***	-0.194***	1.118***
Nobs.	10799	7506	7384

Notes: The table shows means by gender. Standard deviations in parenthesis. Mean differences significant at the * 10%, ** 5 % and *** 1% levels.

Source: NCDS Longitudinal Data.

Table 3. Impact of puberty development on Math outcomes by gender. Family, Birth, School, and Physical Controls. Contemporaneous Model

	(1)	(2)	(3)	(4)
	At age 11		At age 16	
	Boys	Girls	Boys	Girls
Puberty Index at t	0.207**	0.173***	0.220***	-0.033
	(0.082)	(0.056)	(0.056)	(0.063)
Obs.	5501	5283	3186	3623

Notes. Standard errors are in parentheses. The estimates are obtained from a contemporaneous specification estimated at each child age. Each regression includes main controls (white race, mother's age at birth, whether the mother is a lone parent, and whether the mother has post-secondary education) and birth controls (birth order, birth weight, term pregnancy, and birth complications), other family controls (English non-spoken at home, maternal smoking during pregnancy, maternal working before birth, and mother's height), and physical controls (mother's BMI, and child's BMI at 7). *Significant at 5% level; **Significant at 1% level.

Source: NCDS Longitudinal Data.

Table 4. Impact of puberty development on Math outcomes by gender. Family, Birth, School, and Physical Controls. Value-Added and Cumulative Value-Added Models

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	At 11		At 16					
	Value Added		Value Added		Cum. Value Added		Cum. Value Added-IV	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Puberty Index at t	0.115	0.095*	0.083*	-0.095**	0.078*	-0.117**	0.065	-0.140**
	(0.070)	(0.049)	(0.043)	(0.048)	(0.046)	(0.052)	(0.048)	(0.056)
Puberty Index at t-1					-0.027	0.147***	-0.029	0.135**
					(0.073)	(0.052)	(0.077)	(0.057)
Math scores at t-1	0.468***	0.431***	0.674***	0.616***	0.669***	0.620***	0.780***	0.824***
	(0.011)	(0.011)	(0.012)	(0.013)	(0.013)	(0.014)	(0.027)	(0.031)
Obs.	5501	5283	3186	3623	3186	3623	3307	3158

Notes. Standard errors are in parentheses. The estimates are obtained from the value added specifications of Equations 2 to 4 at each child's age. Each regression includes main controls (white race, mother's age at birth, whether the mother is a lone parent, and whether the mother has post-secondary education) and birth controls (birth order, birth weight, term pregnancy, and birth complications), other family controls (English non-spoken at home, maternal smoking during pregnancy, maternal working before birth, and mother's height), and physical controls (mother's BMI, and child's BMI at 7). *Significant at 5% level; **Significant at 1% level.

Source: NCDS Longitudinal Data.

Table 5. Impact of puberty development on the Math gender gap. Value-Added specifications.

	At 11	At 16		
	(1)	(2)	(3)	(4)
	Value Added	Value Added	CVA	CVA-IV
Panel A. No interactions				
Female	0.002	-0.179***	-0.179***	-0.171***
	(0.014)	(0.014)	(0.014)	(0.015)
Panel B. Puberty Development				
Female	-0.005	-0.055	-0.081	-0.068
	(0.020)	(0.048)	(0.051)	(0.054)
Puberty Development t	0.104	0.090**	0.081*	0.06
	(0.069)	(0.042)	(0.045)	(0.048)
Puberty Development t * Female	-0.011	-0.179***	-0.198***	-0.195***
	(0.083)	(0.063)	(0.069)	(0.072)
Puberty Development t-1			-0.029	-0.03
			(0.072)	(0.077)
Puberty Development t-1 *Female			0.183**	0.172*
			(0.088)	(0.094)
Observations	10793	7727	7727	6465

Notes. Standard errors are in parentheses. Each regression includes main controls (white race, mother's age at birth, whether the mother is a lone parent, and whether the mother has post-secondary education) and birth controls (birth order, birth weight, term pregnancy, and birth complications), other family controls (English non-spoken at home, maternal smoking during pregnancy, maternal working before birth, and mother's height), and physical controls (mother's BMI, and child's BMI at 7). *Significant at 5% level; **Significant at 1% level. Source: NCDS Longitudinal Data.

Table 6. Impact of puberty development on Math outcomes by gender. Family, Birth, School, and Physical Controls. Basic, Two-Step and Latent Factors Cumulative Value-Added Models. Age 16

	(1)	(2)	(3)	(4)	(5)	(6)
	CVA-IV		Two-Step CVA		Latent Factors CVA	
	Boys	Girls	Boys	Girls	Boys	Girls
Panel A:						
Puberty Index at t	0.065	-0.140**	0.117**	-0.106*	0.072	-0.205***
	(0.048)	(0.056)	(0.049)	(0.055)	(0.049)	(0.073)
Puberty Index at t-1	-0.029	0.135**	-0.019	0.151***	-0.041	0.212***
	(0.077)	(0.057)	(0.078)	(0.056)	(0.079)	(0.074)
Math scores at t-1	0.780***	0.824***			0.736***	1.233***
	(0.027)	(0.031)			(0.095)	(0.139)
Reading scores at t-1					0.064	-0.533***
					(0.095)	(0.143)
Non-cognitive scores at t-1					-0.023	-0.002
					(0.027)	(0.035)
Panel B:						
Age at development	0.268*	-0.195	0.367**	-0.085	0.342**	-0.209
	(0.154)	(0.133)	(0.152)	(0.132)	(0.160)	(0.170)
Age at development squared	-0.010*	0.008	-0.014**	0.003	-0.013**	0.008
	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.007)
Math scores at t-1	0.748***	0.790***			0.586***	1.190***
	(0.029)	(0.029)			(0.101)	(0.132)
Reading scores at t-1					0.188*	-0.509***
					(0.101)	(0.136)
Non-cognitive scores at t-1					-0.006	0.028
					(0.029)	(0.034)
Obs.	3307	3158	3307	3158	3307	3158

Notes. Standard errors are in parentheses. The estimates are obtained from value added specifications of Equations 4 to 6 at each child's age. Each regression includes main controls (white race, mother's age at birth, whether the mother is a lone parent, and whether the mother has post-secondary education) and birth controls (birth order, birth weight, term pregnancy, and birth complications), other family controls (English non-spoken at home, maternal smoking during pregnancy, maternal working before birth, and mother's height), and physical controls (mother's BMI, and child's BMI at 7). *Significant at 5% level; **Significant at 1% level.

Source: NCDS Longitudinal Data.

Table 7. Validation Exercise

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	At 11		At 16					
	Value Added		Value Added		Cum. Value Added		Cum. Value Added-IV	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Panel A: Dep. Var. Birth Weight								
Puberty Index at t	0.113	-0.076	-0.03	-0.054	-0.051	-0.086	-0.053	-0.111
	(0.093)	(0.064)	(0.067)	(0.077)	(0.072)	(0.083)	(0.077)	(0.086)
Puberty Index at t-1					0.221*	-0.061	0.224*	-0.08
					(0.115)	(0.078)	(0.122)	(0.082)
Panel B: Dep. Var. Mother with Post-Secondary Education								
Puberty Index at t	-0.037	-0.021	-0.047*	0.02	-0.055*	0.018	-0.066**	0.023
	(0.036)	(0.026)	(0.026)	(0.031)	(0.028)	(0.034)	(0.029)	(0.035)
Puberty Index at t-1					0.007	-0.03	0.03	-0.027
					(0.046)	(0.034)	(0.049)	(0.036)
Panel C: Dep. Var. Mother Smoked during Pregnancy								
Puberty Index at t	-0.021	-0.011	-0.017	0.063*	-0.014	0.022	-0.01	0.037
	(0.044)	(0.031)	(0.031)	(0.037)	(0.033)	(0.040)	(0.035)	(0.042)
Puberty Index at t-1					-0.018	0.003	-0.021	-0.009
					(0.054)	(0.038)	(0.057)	(0.040)
N	5501	5283	3938	3789	3610	3425	3307	3158

Notes. Standard errors are in parentheses. The estimates are obtained from the value-added specifications of Equations 2 to 4 at each child's age. For each dependent variable, the rest of covariates in Table B.2 are included as controls. These include main controls (white race, mother's age at birth, whether the mother is a lone parent, and whether the mother has post-secondary education) and birth controls (birth order, birth weight, term pregnancy, and birth complications), other family controls (English non-spoken at home, maternal smoking during pregnancy, maternal working before birth, and mother's height), and physical controls (mother's BMI, and child's BMI at 7). *Significant at 5% level; **Significant at 1% level.

Source: NCDS Longitudinal Data.

Table 8: Heterogeneity in the impact of puberty development by maternal education, initial skills and self-assessed math efficacy.

	(1)	(2)	(3)	(4)	(5)	(6)
	At 11		At 16			
	Value Added		Cum. Value Added		Cum. Value Added-IV	
	Boys	Girls	Boys	Girls	Boys	Girls
Panel A: By Initial Skills						
Top Half at Age 7	0.082*	-0.001	0.054	0.127	-0.076	-0.066
	(0.043)	(0.044)	(0.063)	(0.089)	(0.078)	(0.099)
Puberty Index at t	0.118	0.149**	0.043	-0.138**	0.019	-0.181***
	(0.093)	(0.062)	(0.056)	(0.062)	(0.063)	(0.069)
Puberty Index at t-1			0.058	0.196***	0.081	0.178**
			(0.092)	(0.065)	(0.103)	(0.076)
Top Half*Puberty Index at t	-0.006	-0.128	0.077	0.044	0.09	0.09
	(0.139)	(0.094)	(0.093)	(0.110)	(0.097)	(0.113)
Top Half*Puberty Index at t-1			-0.203	-0.105	-0.231	-0.1
			(0.148)	(0.103)	(0.156)	(0.110)
Panel B: By Self-Assessed Math Performance						
Above Average Self-Assessed Performance	0.534***	0.485***	0.394***	0.616***	0.308***	0.423***
	(0.038)	(0.050)	(0.074)	(0.122)	(0.078)	(0.126)
Puberty Index at t	0.151**	0.072	0.042	-0.119**	0.031	-0.151***
	(0.075)	(0.050)	(0.049)	(0.053)	(0.052)	(0.057)
Puberty Index at t-1			0.041	0.155***	0.028	0.142**
			(0.077)	(0.054)	(0.082)	(0.059)
Avobe Average*Puberty Index at t	-0.065	0.137	0.185*	-0.019	0.193*	0.047
	(0.179)	(0.139)	(0.109)	(0.151)	(0.112)	(0.151)
Avobe Average*Puberty Index at t-1			-0.18	-0.1	-0.137	-0.09
			(0.175)	(0.141)	(0.181)	(0.142)
N	5501	5283	3610	3425	3307	3158

Notes. Standard errors are in parentheses. The estimates are obtained from the value added specifications of Equations 2 to 4 at each child's age. Each regression includes main controls (white race, mother's age at birth, whether the mother is a lone parent, and whether the mother has post-secondary education) and birth controls (birth order, birth weight, term pregnancy, and birth complications), other family controls (English non-spoken at home, maternal smoking during pregnancy, maternal working before birth, and mother's height), and physical controls (mother's BMI, and child's BMI at 7). *Significant at 5% level; **Significant at 1% level.

Source: NCDS Longitudinal Data.

Table 9. Impact of puberty development on English outcomes by gender. Family, Birth, School, and Physical Controls. Ages 11 and 16

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	At 11		At 16					
	Value Added		Value Added		Cum. Value Added		Cum. Value Added-IV	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Panel A:								
Puberty Index at t	0.103	0.144***	0.211***	0.117***	0.198***	0.100**	0.127***	0.125**
	(0.069)	(0.047)	(0.039)	(0.043)	(0.043)	(0.046)	(0.046)	(0.051)
Puberty Index at t-1					-0.061	0.01	-0.078	-0.067
					(0.070)	(0.042)	(0.077)	(0.048)
Math scores at t-1	0.525***	0.500***	0.672***	0.681***	0.675***	0.676***	0.939***	0.965***
	(0.011)	(0.012)	(0.012)	(0.012)	(0.013)	(0.013)	(0.027)	(0.031)
Obs.	5501	5283	3186	3623	3186	3623	3307	3158

Notes. Standard errors are in parentheses. The estimates are obtained from different specifications. Each regression includes main controls (white race, mother's age at birth, whether the mother is a lone parent, and whether the mother has post-secondary education) and birth controls (birth order, birth weight, term pregnancy, and birth complications), other family controls (English non-spoken at home, maternal smoking during pregnancy, maternal working before birth, and mother's height), and physical controls (mother's BMI, and child's BMI at 7). *Significant at 5% level; **Significant at 1% level.

Source: NCDS Longitudinal Data.

Appendix A. Appendix Figures: NCDS

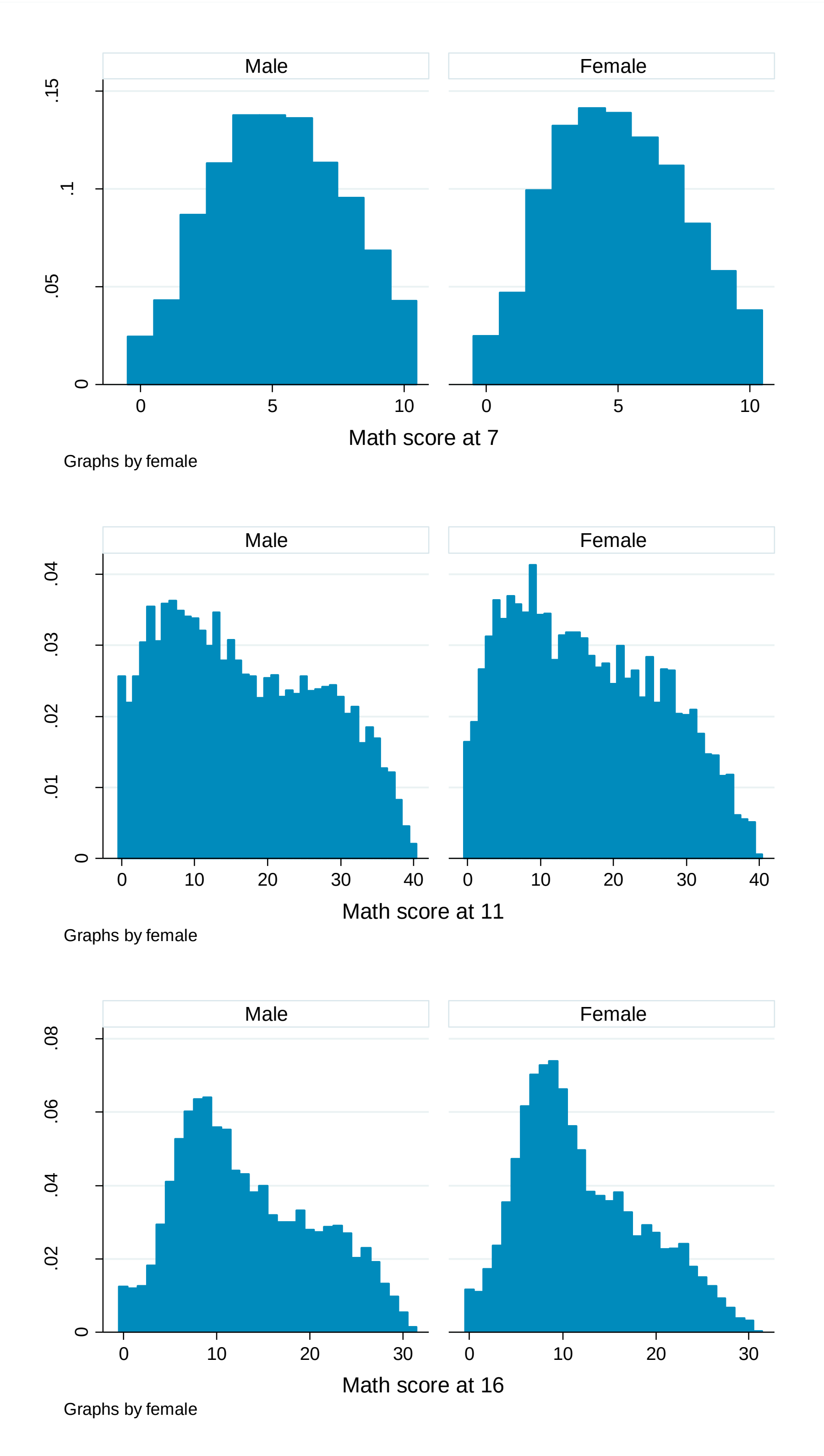


Figure A1. Math scores by gender over the life course

Source: NCDS58

Appendix B. Appendix Tables

Table B.1 Determinants of Sample Attrition

	Female Sample		Male Sample	
	(1)	(2)	(3)	(4)
	Age 11	Age16	Age 11	Age16
Other ethinic group	0.037*	0.065*	-0.035*	0.086**
	(0.020)	(0.037)	(0.020)	(0.035)
Other ethinic group missing	0.916***	0.596***	0.909***	0.584***
	(0.005)	(0.010)	(0.005)	(0.009)
Mother's age at birth	-0.001**	0.000	-0.000	0.001*
	(0.000)	(0.001)	(0.000)	(0.001)
Mother's age at birth missing	-0.009	-0.164**	0.003	-0.018
	(0.042)	(0.076)	(0.043)	(0.077)
Mother's sole parent	0.012	-0.001	-0.005	0.030
	(0.012)	(0.023)	(0.013)	(0.023)
Mother's sole parent missing	0.111	0.002	-0.026	-0.144
	(0.092)	(0.168)	(0.134)	(0.237)
Mother post secondary education	0.004	-0.049***	0.011*	-0.037***
	(0.006)	(0.010)	(0.006)	(0.010)
Mother post secondary education missing	0.012	0.182**	0.008	0.016
	(0.041)	(0.075)	(0.042)	(0.075)
Constant	0.096***	0.392***	0.078***	0.376***
	(0.012)	(0.022)	(0.012)	(0.022)
N. Obs.	8,957	8,957	9,600	9,600

Notes. Standard errors are in parentheses. Each column in each panel estimates a linear probability model of the likelihood of having missing information in the 7, 11, 14 or 16 year-old samples as indicated in the column header. *Significant at 5% level; **Significant at 1% level.

Source: NCDS58 and ALSPAC90 Longitudinal Data.

Table B2. Descriptive statistics: Control variables. NCDS data.

	NCDS58		
	Female	**Male**	**Difference**
Panel A. Maternal Demographic Controls			
Other ethinic group	0.0153	0.0157	-0.000
	(0.123)	(0.124)	
Mother's age at birth	27.5210	27.4282	0.070
	(5.560)	(5.512)	
Mother's sole parent	0.0334	0.0294	0.005**
	(0.180)	(0.169)	
Mother post secondary education	0.2446	0.2373	0.003
	(0.430)	(0.425)	
Panel B. Birth and Pregnancy Controls			
Birth order	2.2480	2.2416	0.001
	(1.533)	(1.519)	
Birth weight	7.1831	7.4619	-0.268***
	(1.085)	(1.112)	
Term pregnancy	40.1530	40.0848	0.060***
	(1.617)	(1.633)	
Birth complications	0.0889	0.0987	-0.011***
	(0.285)	(0.298)	
Maternal smoking during pregnancy	0.3914	0.3745	0.012**
	(0.488)	(0.484)	
Maternal working before birth	0.3612	0.3652	-0.003
	(0.480)	(0.481)	
Panel C. Maternal Controls			
English non-spoken at home	0.0959	0.0947	-0.001
	(0.295)	(0.293)	
Mother’s height	63.2948	63.3030	0.023
	(2.420)	(2.438)	
Mother's BMI	0.0049	0.0049	-0.000**
	(0.001)	(0.001)	
Panel D. School Controls			
Class size	31.9303	31.6675	0.281**
	(9.176)	(9.220)	
Type of school-compehensive	0.4987	0.5010	0.006
	(0.500)	(0.500)	
Type of school-grammar	0.1095	0.0891	0.017***
	(0.312)	(0.285)	
Type of school-modern	0.1804	0.1835	0.002
	(0.385)	(0.387)	
Type of school-private	0.0457	0.0492	-0.002
	(0.209)	(0.216)	
Panel E. Physical Controls			
Child's BMI at 7	15.7624	15.8442	-0.054***
	(1.790)	(1.548)	

Notes: The table shows means by gender. Standard deviations in parenthesis. Mean differences significant at the * 10%, ** 5 % and *** 1% levels.

Source: NCDS Longitudinal Data.

Table B3. Descriptive statistics: Potential mechanisms variables. NCDS data.

	At age 11			At age 16		
	Female	Male	Difference	Female	Male	Difference
Panel A. Family Characteristics						
Mother employed	0.5648	0.5399	0.019***	0.5083	0.5179	0.004
	(0.496)	(0.498)		(0.500)	(0.500)	
Mother's occupation manager	0.0815	0.0711	0.008***	0.0933	0.0988	-0.001
	(0.274)	(0.257)		(0.291)	(0.298)	
Mother expects child manager				0.2593	0.1874	0.061***
				(0.438)	(0.390)	
Panel B. School Characteristics						
Female teacher	0.4382	0.3909	0.045***			
	(0.496)	(0.488)				
Co-educational school				0.6270	0.6474	-0.006
				(0.484)	(0.478)	
Number of Math subjects				1.0230	1.0703	-0.040***
				(0.326)	(0.340)	
Number of English subjects				1.3122	1.2210	0.072***
				(0.493)	(0.454)	
Teacher expects child in STEM				0.0502	0.3060	-0.202***
				(0.218)	(0.461)	
Panel C. Own Child Features						
Overweight child	0.1372	0.1252	0.013***	0.0974	0.0692	0.021***
	(0.344)	(0.331)		(0.296)	(0.254)	
Above Rutter score cutoff	0.1376	0.1763	-0.030***	0.1883	0.1758	0.010***
	(0.344)	(0.381)		(0.391)	(0.381)	
Above average self-assessed Math				0.0901	0.1740	-0.065***
Above average self-assessed English				0.2223	0.1915	0.022***
				(0.416)	(0.393)	

Notes: The table shows means by gender. Standard deviations in parenthesis. Mean differences significant at the * 10%, ** 5 % and *** 1% levels.

Source: NCDS Longitudinal Data.

Table B.4. Family Mechanisms. Impact of puberty development on Math outcomes by gender. Value Added (age 11) Cumulative Value Added Instrumental Variables Model (age16)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	11 years of age						16 years of age							
	Benchmark		Mother employed		Mother manager		Benchmark		Mother employed		Mother manager		Expects child manager	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Panel A:														
Puberty Index at t	0.115	0.095*	0.115	0.097**	0.114	0.093*	0.065	-0.140**	0.068	-0.136**	0.067	-0.140**	0.071	-0.140**
	(0.070)	(0.049)	(0.070)	(0.049)	(0.070)	(0.048)	(0.048)	(0.056)	(0.048)	(0.056)	(0.048)	(0.056)	(0.048)	(0.055)
Family variable at t			-0.03	-0.03	0.008	0.092**			-0.006	-0.035	-0.074*	0.008	0.234***	0.131***
			(0.021)	(0.021)	(0.041)	(0.037)			(0.028)	(0.027)	(0.039)	(0.044)	(0.032)	(0.026)
Puberty Index at t-1							-0.029	0.135**	-0.033	0.144**	-0.031	0.135**	-0.046	0.126**
							(0.077)	(0.057)	(0.077)	(0.057)	(0.077)	(0.057)	(0.075)	(0.057)
Family variab at t-1									-0.051**	-0.066***	0.013	-0.006		
									(0.025)	(0.025)	(0.045)	(0.047)		
Obs.	5501	5283	5501	5283	5501	5283	3307	3158	3307	3158	3307	3158	3307	3158

Notes. Standard errors are in parentheses. The estimates are obtained from a VA specification in Columns 1 to 6 and a CVA-IV specification in Columns 7 to 14. Each regression includes main controls (white race, mother's age at birth, whether the mother is a lone parent, and whether the mother has post-secondary education) and birth controls (birth order, birth weight, term pregnancy, and birth complications), other family controls (English non-spoken at home, maternal smoking during pregnancy, maternal working before birth, and mother’s height), and physical controls (mother's BMI, and child's BMI at 7). *Significant at 5% level; **Significant at 1% level.

Source: NCDS Longitudinal Data.

Table B.5. School Mechanisms. Impact of puberty development on Math outcomes by gender. Value Added (age 11) Cumulative Value Added Instrumental Variables Model (age16)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	11 years of age				16 years of age									
	Benchmark		Female teacher		Benchmark		Female teacher		Teacher expects STEM		No. Math subjects		Co-educational school	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Panel A:														
Puberty Index at t	0.115	0.095*	0.117*	0.093*	0.065	-0.140**	0.066	-0.139**	0.064	-0.135**	0.073	-0.141**	0.066	-0.142**
	(0.070)	(0.049)	(0.070)	(0.048)	(0.048)	(0.056)	(0.048)	(0.056)	(0.048)	(0.056)	(0.048)	(0.055)	(0.048)	(0.056)
School variable at t			0.004	-0.003					0.052**	0.249***	0.267***	0.258***	-0.004	-0.041
			(0.021)	(0.021)					(0.024)	(0.048)	(0.034)	(0.040)	(0.032)	(0.031)
Puberty Index at t-1					-0.029	0.135**	-0.03	0.134**	-0.034	0.135**	-0.018	0.125**	-0.03	0.135**
					(0.077)	(0.057)	(0.077)	(0.057)	(0.077)	(0.057)	(0.076)	(0.056)	(0.077)	(0.057)
School variable at t-1							0.022	0.021						
							(0.024)	(0.024)						
Obs.	5283	5501	5283	5501	3307	3158	3307	3158	3307	3158	3307	3158	3307	3158

Notes. Standard errors are in parentheses. The estimates are obtained from a VA specification in Columns 1 to 4 and a CVA-IV specification in Columns 5 to 14. Each regression includes main controls (white race, mother's age at birth, whether the mother is a lone parent, and whether the mother has post-secondary education) and birth controls (birth order, birth weight, term pregnancy, and birth complications), other family controls (English non-spoken at home, maternal smoking during pregnancy, maternal working before birth, and mother’s height), and physical controls (mother's BMI, and child's BMI at 7). *Significant at 5% level; **Significant at 1% level.

Source: NCDS Longitudinal Data.

Table B.6. Own-Child Mechanisms. Impact of puberty development on Math outcomes by gender. Value Added (age 11) Cumulative Value Added Instrumental Variables Model (age16)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	11 years of age						16 years of age					
	Benchmark		Child BMI		Behav. Problems		Benchmark		Child BMI		Behav. Problems	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Panel A:												
Puberty Index at t	0.115	0.095*	0.11	0.094*	0.115	0.085*	0.065	-0.140**	0.077	-0.134**	0.065	-0.142**
	(0.070)	(0.049)	(0.071)	(0.050)	(0.070)	(0.048)	(0.048)	(0.056)	(0.050)	(0.056)	(0.048)	(0.056)
Own-child variable at t			0.008	0	-0.019***	-0.033***			-0.011*	-0.002	-0.003	-0.008**
			(0.005)	(0.005)	(0.003)	(0.003)			(0.006)	(0.005)	(0.003)	(0.004)
Puberty Index at t-1							-0.029	0.135**	-0.026	0.159***	-0.033	0.130**
							(0.077)	(0.057)	(0.077)	(0.059)	(0.077)	(0.057)
Own-child variable at t-1									0.012	-0.009	0.002	0.003
									(0.007)	(0.007)	(0.003)	(0.004)
Obs.	5283	5501	5283	5501	5283	5501	3307	3158	3307	3158	3307	3158

Notes. Standard errors are in parentheses. The estimates are obtained from a VA specification in Columns 1 to 4 and a CVA-IV specification in Columns 5 to 14. Each regression includes main controls (white race, mother's age at birth, whether the mother is a lone parent, and whether the mother has post-secondary education) and birth controls (birth order, birth weight, term pregnancy, and birth complications), other family controls (English non-spoken at home, maternal smoking during pregnancy, maternal working before birth, and mother’s height), and physical controls (mother's BMI, and child's BMI at 7). *Significant at 5% level; **Significant at 1% level.

Source: NCDS Longitudinal Data.

Appendix C. Examples of tests' questions in NCDS

Math questions for age 7 tests

- Peter had 4 toy cars and be bought 2 more. How many toy cars did he have altogether ~

- How many socks are there in 4 pairs?

- What is half of 38?

- A boy spent 4d. a day for 5 days. How much would he have left out of 2 pounds.?

Math questions for age 11 tests

- In a class of -40 pupils 3/4 are girls. How many of the pupils are boys?

- A rectangle whose length is 6 in. and breadth is 4 in. has an area of 24 sq. in. Give the length and breadth of another rectangle whose area is 24, sq. in.

Math questions for age 16 tests

- The solution of the equations $x + y = 8$ and $x - y = 4$ is:

a. $x = 4$ $y = 4$

b. $x = 7$ $y = 3$

c. $x = 5$ $y = 3$

d. $x = 2$ $y = -2$

e. $x = 6$ $y = 2$

- Which of the following is an even number for any whole number of n

a. n2

b. 2n

c. 2n - 1

d. 2n + 1

e. n2 – 2

Reading questions for age 7 tests

- "Point to the first picture. the picture of an elephant. Look at the words in the box beside it. Find the one that says "elephant'. Don't tell anybody which it is. When you have found the word that says 'elephant' draw a ring round it. Go on, do it by yourselves. Find the word that says 'elephant' and draw a ring round it."

Reading question for age 11 tests

- Read the next line carefully and then look up. The line reads, 'A birds lays its eggs in a (pond, stream, cloud, house, nest)’. Underline the correct word.

Reading question for age 16 tests:

- Now go through the sentences below and put a line under the right word in each one. ‘She had been dieting for a month but her weight had not (shown. increased, shrunk, decreased, grown)’.