



Adolescent development and the math gender gap

Cristina Borra^{a,*}, Maria Iacovou^b, Almudena Sevilla^c

^a Universidad de Sevilla and IZA, Spain

^b University of Cambridge, United Kingdom

^c London School of Economics and IZA, United Kingdom

ARTICLE INFO

JEL:

I21

I24

J16

Keywords:

Adolescent development

Educational outcomes

Gender gap in mathematics

ABSTRACT

This paper investigates the determinants of the gap between girls and boys in mathematics performance (the ‘math gap’). We are particularly interested in the role played by pubertal development in explaining the widening of the math gap over adolescence. We estimate rich production function models of math skills, using data from the 1958 British National Child Development Study (NCDS), a longitudinal survey of all British children born in a single week in March 1958 which contains unique information on pubertal development and educational outcomes. Using (cumulative) value-added models, we show that the impact of puberty varies by the age and gender of the child, and that this heterogeneity can explain about two thirds of the math gap that emerges between the ages of 11 and 16. We find also that the widening of the math gap during adolescence is driven by markers of pubertal development which are publicly evident, rather than by markers which are only privately evident; and that the relationship between puberty and math scores is strongly associated with children’s self-perceived ability in math. Taken together, these results suggest that the mechanisms underlying the relationship between pubertal development and the math gender gap are socially rather than biologically driven.

1. Introduction

The gender gap in academic achievement has reversed in recent decades, with more women than men now holding college degrees (Ceci et al., 2014; Goldin et al., 2006; OECD, 2019a). However, research has consistently shown that in many countries, boys still tend to outperform girls in mathematics (Guiso et al., 2008; Kane and Mertz, 2012). The ‘math gap’ is present among children of primary school age, and appears to widen as they progress from primary to secondary school: on average, girls’ math scores are about 9% lower than those of boys at age 15.¹ This gender gap in math is a factor contributing to the underrepresentation of women in STEM fields,

* Corresponding author at: Department of Economics and Economic History, Universidad de Sevilla, Ramon y Cajal 1, Seville 41018, Spain.

E-mail address: cborra@us.es (C. Borra).

¹ Studies from various countries (Bharadwaj et al., 2016; Borgonovi et al., 2021; Contini et al., 2017; Ellison and Swanson, 2021) document a math gender gap which approximately triples over adolescence, from about 3% of a standard deviation at age 9/10 to about 9% at age 15/16. Studies using the international TIMMS and PISA data sets (Mullis et al., 2016; OECD, 2019b, 2014, 2010) confirm that the widening of the math gender gap over adolescence is a worldwide phenomenon. Our own analysis of UK cohorts (the 1993 cohort, interviewed aged 9-10 in TIMMS 2003 and aged 15-16 in PISA 2009; the 1996-97 cohort, interviewed aged 9-10 in TIMMS 2007 and aged 15-16 in PISA 2012; and the 2001-2002 cohort, interviewed aged 9-10 in TIMMS 2011 and aged 15-16 in PISA 2018) indicates no significant math gender gap at age 9, widening to a significant positive gap in favour of boys at age 15-16. The size of the math gap seems to have diminished over time from about 21% of a standard deviation in 2009 to about 12% of a standard deviation in 2012 and 2018, but remains sizeable and significant.

<https://doi.org/10.1016/j.eurocorev.2023.104542>

Received 11 May 2022; Received in revised form 12 July 2023; Accepted 17 July 2023

Available online 21 July 2023

0014-2921/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

which tend to be high-paying (Black et al., 2008; Blau and Kahn, 2017; Card and Payne, 2021; Joensen and Nielsen, 2016; Kahn and Ginther, 2018). Understanding the mechanisms that influence math performance among boys and girls is thus crucial for promoting equal opportunities and gender equality. This paper explores whether pubertal development can explain the different trends observed in boys' and girls' math performance over the adolescent period, and investigates potential underlying mechanisms.

We estimate rich production function models of boys' and girls' math skills at ages 11 and 16, incorporating measures of pubertal development at those ages as well as controls for other school and family inputs. We use data from the first four waves of the 1958 National Child Development Study (NCDS), collected during pregnancy and birth, and then at 7, 11 and 16 years old. The NCDS is particularly useful for several reasons: its wide coverage (which encompasses detailed data on family background, schooling, academic and cognitive skills, and pubertal development); its longitudinal design, which means that repeated measures of all the variables of interest are available; and the fact that multiple indicators of pubertal development were collected, including standardized assessments by trained health personnel.

We pay special attention to the assumptions needed to identify key parameters in the models we estimate (Del Boca et al., 2017; Fiorini and Keane, 2014; Todd and Wolpin, 2003). Estimating child production functions is a difficult undertaking, given the endogeneity and imperfect observability of essentially all inputs. Our approach is to implement several different empirical strategies based on previous research in the area. We address two key challenges for the correct identification of parameters. First, we seek to minimize selection bias due to unobservable factors. Measures of pubertal development may be correlated with omitted variables, such as nutrition or genetic endowments. Our longitudinal data allow us to use child fixed-effects and value-added models, which coupled with the wide range of additional control variables available allows us to account for unobservable factors. Second, we minimize the potential attenuation bias induced by measurement error in indicators of lagged achievement (Andrabi et al., 2011). In our case, if past achievement in math is incorrectly measured, for example because a child performs poorly due to feeling unwell on the day of the test, the estimated coefficient on past performance may be biased downwards, which could in turn bias the estimated impact of pubertal development. Our rich data set containing math scores measured at multiple points in a child's life allows us to address measurement error in lagged measures of performance through an instrumental-variables strategy (Del Bono et al., 2016).

Our findings are generally consistent across models and suggest that pubertal development plays a substantial role in explaining the widening of the math gender gap over the adolescent years. We find that puberty affects math performance differently for boys and girls. For girls, more advanced pubertal development is associated with better performance in math at age 11, but with poorer performance at age 16. For boys, pubertal development is not associated with math scores at age 11, but is associated with higher scores at age 16. Direct estimates of the size of the math gap reveal that, once we control for prior achievement in math, no math gender gap remains at age 11, but a gap of about 18% of a standard deviation is observed at age 16. After controlling for pubertal development, the estimated math gap at age 16 falls to about 8% of a standard deviation; thus, well over half of the math gender gap at age 16 may be explained by pubertal development. Our results are robust to a range of tests for the presence of unobserved heterogeneity and deviations from standard model assumptions.

We then discuss possible mechanisms driving the relationship between pubertal development and the math gap. There are plausible arguments for both biological and social mechanisms. We investigate the relative impacts of markers of puberty that are evident in public (eg. breast development) and those that are not publicly evident (eg. pubic hair growth); our finding that the impact of pubertal development on math scores is driven by changes in publicly evident markers of puberty suggests a socially-driven mechanism. This is also supported by our finding of a substantially larger impact of pubertal development in boys with higher self-perception of their own ability in math, in the context of evidence that self-assessed ability is largely driven by social stereotypes (Bordalo et al., 2016; Carlana, 2019; Coffman, 2014).

This paper contributes to a growing literature in economics concerned with adolescent development and its relationship to socioeconomic outcomes (Del Boca et al., 2019; Doepke et al., 2019).² To our knowledge, this is the first paper to provide evidence of the impact of adolescent development on mathematical performance.³ In contrast to previous correlational studies that look at the contemporaneous influence of adolescent development on educational performance (Cavanagh et al., 2007; Koerselman and Pekkarinen, 2018; Martin and Steinbeck, 2017), the availability of rich longitudinal data and detailed medical assessments of pubertal development enables us to estimate dynamic skill production function models of math performance by gender and age.

We also contribute to the literature on gender gaps. Previous studies have considered a multitude of factors for explaining the math gap, including the gender composition of the child's class (Booth et al., 2018; Cools et al., 2022); the number of math subjects and math hours taken by students (Ellison and Swanson, 2021; Machin and McNally, 2005); maternal employment and occupation (Carneiro et al., 2013; Fryer and Levitt, 2010); and teachers' and parents' expectations and attitudes (Alan et al., 2018; Carlana, 2019; Dossi et al., 2021; Fryer and Levitt, 2010; Iriberry and Rey-Biel, 2019; Lavy and Sand, 2018; Maloney et al., 2015). In this paper we have shown that pubertal development is key to the understanding of the evolution of the math gender gap, and that this relationship is

² This literature has looked at risky behaviour (Aizer, 2017; Rodríguez-Planas et al., 2022), time investment behaviour (del Boca et al. 2017), non-cognitive skills (Lei and Lundberg, 2020; Schurer, 2017), and the impact of educational interventions during adolescence (Guryan et al., 2023) among others.

³ There is considerable debate in the literature on whether it makes sense to talk about the causal effect of a non-manipulable factor (see Holland, 1986 and Heckman and Pinto, 2015 among others). However, as highlighted in the paper, the gender gap in math can be partly attributed to the timing of pubertal development and the resulting gender roles. Therefore, in line with Greiner and Rubin (2011), even though pubertal development itself may not be manipulable, the potential impact of its timing on gender roles and subsequent math performance may be subject to intervention and can be studied causally.

driven by social factors and social stereotypes.

The paper is organized as follows. [Section 2](#) describes the data used. [Section 3](#) presents the different specifications of the math skill production function which are used for empirical estimation. [Section 4](#) presents our main estimation results together with specification and identification checks. [Section 5](#) discusses potential mechanisms underlying the relationships which we uncover. [Section 6](#) concludes.

2. Data

The 1958 National Child Development Study (NCDS) is a longitudinal survey that sampled all children born in a single week of 1958. Originally designed as a perinatal mortality study, the initial birth survey was followed by sweeps carried out when cohort members were 7, 11, 16, 23, 33, 41, 46, 50, and 55 years old (see [Centre for Longitudinal Studies, 2023](#)). NCDS data have formed the basis for a large number of studies, for example those by [Persico et al. \(2004\)](#) and [Case and Paxson \(2008\)](#) studying the relationship between adolescent height and labour market outcomes, and by [Goodman \(2014\)](#) examining the association between handedness and human capital accumulation.

The surveys carried out at ages 7, 11 and 16 collected information on cognitive outcomes, schooling, health, and family circumstances. The NCDS has special features that make it particularly suitable for the study of the relationship between puberty and educational achievement. First, thanks to the longitudinal nature of the survey, we are able to construct different measures of pubertal development for ages 11 and 16, coincident with the measurement of cognitive achievement at those ages. Second, the NCDS offers extremely rich information on the child's family and school environments. And third, the study offers both self-reported and medical assessments of pubertal 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 Millennium 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 only measures of pubertal development assessed by parents or children themselves; these are generally considered to be less accurate than objective assessments by doctors ([Brooks-Gunn et al., 1987](#); [Walker et al., 2020](#)). Additionally, the fact that all members of the NCDS cohort were born within a single week means that issues arising from differences in the chronological age at which their development was assessed are minimized. And finally, because examinations of pubertal development were performed together with the other medical examinations carried out in the NCDS, sample attrition was minimized.

The original (pregnancy and birth) sweep of NCDS included 8957 girls and 9600 boys. As with any longitudinal survey, attrition means that follow-up samples are smaller: our main results are based on 5501 boys and 5283 girls (age 11) and 3610 boys and 3425 girls (age 16). Columns 1–4 in [Table A.1](#) ([Appendix A](#)) show the relationship between attrition and selected covariates. We acknowledge that our sample may not be adequately representative of the non-white population. However, the small size of the coefficients for the remainder of the independent variables suggests that selection on observables is quantitatively weak. In order to maximize sample size we use as many observations as are available for each regression; sample size therefore varies between models.⁴ In [Section 4](#) below we perform a robustness test, showing that our results are virtually the same, albeit slightly less efficiently estimated, when using a consistent sample at age 11.

2.1. Test scores

Our main dependent variables are standardized math scores. The NCDS administered math tests at ages 7, 11 and 16. At age 7, 10 problems were read to the child, and children's scores ranged from 0 to 10. At age 11, a written test contained 40 questions, with scores ranging from 0 to 40; at age 16, the test contained 31 questions, and scores ranged from 0 to 31. Examples of the problems in each test may be found in [Appendix B](#). All scores are standardized to have a mean of 0 and a standard deviation of 1 across the entire cohort for the year in which they were observed, so that all coefficients can be read in units of one standard deviation.⁵ Although all the children in the sample were born in a single week in 1958, the tests were taken at different times; in the empirical analyses we include controls for the child's age at testing, to allow for comparisons between children of different ages.

[Fig. 1](#), and [Panel A in Table 1](#), illustrate that a gender gap in math scores is evident as early as 7 years old, with a difference of around 8% of a standard deviation. This gap decreases to around 5% of a SD at the age of 11, but widens substantially to over 20% of a SD by age 16.

2.2. Measures of adolescent development

Our main explanatory variables are measures of pubertal development, collected at ages 11 and 16, via both medical assessments and self-reports by the child.

⁴ As described in [Sections 3](#) and [4](#), we employ a range of empirical models to estimate math scores at ages 11 and 16. All models, with the exception of a fixed-effects specification, use background data drawn from the original pregnancy and birth sweep. All require contemporaneous measures of math achievement and pubertal development; most also use one-period-lagged measures of achievement and some, which we can estimate only at age 16, use two-period lagged measures.

⁵ [Fig. C.1](#) in [Appendix C](#) shows the distributions of raw math scores by age and gender.

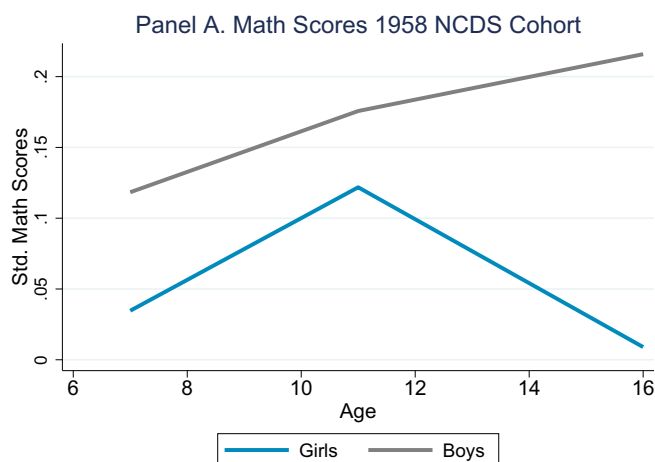


Fig. 1. Standardized math scores by gender at ages 7, 11 and 16.

Note: Math scores were standardized in each wave to have a mean of zero for the whole cohort. The values plotted are for cohort members with complete data at ages 7, 11 and 16; mean values for this sample are higher than zero.

Source: NCDS. 6461 observations.

Table 1

Descriptive statistics: standardized math scores and puberty index.

	Age 7	Age 11	Age 16
Panel A. Standardized Math Scores			
Girls	0.0346 (0.975)	0.1219 (0.963)	0.0090 (0.963)
Boys	0.1184 (0.990)	0.1757 (1.009)	0.2159 (1.031)
Mean gender difference significant	-0.084***	-0.054*	-0.207***
Panel B. Puberty Index			
Girls		0.2327 (0.216)	0.8014 (0.207)
Boys		0.1510 (0.146)	0.6109 (0.240)
Mean gender difference significant		0.082***	0.190***
No. of observations	6461	6461	6461

Notes: The table presents means by age and gender. Math scores were standardized in each wave to have a mean of zero for the whole cohort. The values plotted are for cohort members with complete data at ages 7, 11 and 16; the means shown are therefore a little higher than zero. Standard deviations in parentheses. Gender differences significant at the * 10%, ** 5% and *** 1% levels.

Source: NCDS.

Our principal measure of adolescent development is an index derived from the medical assessments carried out at ages 11 and 16. Trained health personnel assessed pubertal development at ages 11 and 16, using standard procedures (see Brooks-Gunn et al., 1987; Coleman and Coleman, 2002).

At age 11, the NCDS collected data on the development of pubic hair (both girls and boys), breast development (girls) and genitalia (boys). All three measures were coded on a scale from 1 (no development) to 5 (advanced/adult development). Our *age 11 puberty index* is constructed, following (Koerselman and Pekkarinen, 2018), by adding together the two relevant measures for each child, which produces an index ranging from 2 to 10.

At age 16, the survey collected information on the development of pubic and axillary (armpit) hair for both girls and boys, with both measured on a 4-point scale (1 absent; 2 sparse; 3 intermediate; 4 adult). In addition, the survey recorded the growth of facial hair for boys (1 absent; 2 sparse; 3 adult) and breast development for girls (1 absent; 2 intermediate; 3 adult). Adding together the three relevant measures for each child produces our *age 16 puberty index* ranging from 3 to 11.

Both puberty indices were then rescaled to the 0–1 interval, with 0 corresponding to no development on any measure, and 1 representing ‘adult’ development on all measures. These rescaled indices may be interpreted as the proportion of pubertal maturation which the individual had undergone at the time of assessment.

In addition to the measures collected by trained personnel, the survey also includes information reported by cohort members themselves on the age at menarche (girls) and the age at which the voice broke (boys). Both of these constitute distinct events in puberty (Day et al., 2015). Age at menarche, although not adequately validated, has been widely used in epidemiological studies as a proxy for pubertal stage (Walker et al., 2020); the voice breaking is a validated marker of boys’ pubertal development, associated with

Tanner stages 3 and 4 of late puberty (Hodges-Simeon et al., 2013). This variable, which we call *age at development* is used as a supplementary measure of pubertal development.⁶

The NCDS has the advantage that all the cohort members are almost exactly the same age, and (in contrast with some other surveys) the measures of pubertal development are therefore collected at very similar ages. However, they are not collected at identical ages. We therefore control for the calendar month in which assessments were performed, in order to assure comparability between individuals.

Fig. 2 presents the distributions of the puberty index at ages 11 and 16, and of the age at development variable, for boys and girls separately. Panel B in Table 1 also presents descriptive statistics for the puberty index. At age 11, both boys and girls are in the early stages of development; mean values on the 0–1 scale are 0.23 (girls) and 0.15 (boys). For girls, the mean of 0.23 would translate back to a score of 3.86 on the original 2–10 scale – a little below a score of 4, which could arise from a score of 2 (early -to-mid-stage development) on both the breast and pubic hair scales. For boys, the mean of 0.15 would translate back to a score of 3.2 on the original scale (corresponding approximately to, for example, a development level of 2 for genitalia and a level of 1 (no development) for pubic hair). At 16 years of age both boys and girls are in the second half of their pubertal development, with mean scores of 0.80 for girls and 0.61 for boys. These scores translate back to scores of 9.4 and 7.9 respectively on the original 3–11 scale.

Gender disparities are also present in the age at development variable (see Fig. 2), with the menarche occurring on average at age 12 and a half years, and the voice breaking on average at just under 14 years.

2.3. Controls and other variables used in the analysis

The NCDS offers a vast array of potential control variables; we are therefore able to control both for the basic maternal demographics usually included in math gender gap analyses (see for instance Fryer and Levitt, 2010), but also for a range of additional variables which can safely be considered exogenous to the impact of pubertal development on cognitive outcomes. Descriptive statistics for these variables are presented by gender in Table A.2 in Appendix A.

Panel A of Table A.2 contains information on *basic controls* (ethnicity; mother's age at the birth of the cohort child; lone parenthood; and maternal education). We also control for variables relating to *birth and pregnancy* (birth order, birth weight, gestation at birth, and indicators of birth complications, maternal smoking during pregnancy, and whether the mother had a job before birth); *additional family background variables* (English not spoken at home, indicators of the mother's height and body mass index (BMI), and region of residence); characteristics of the *child's school* (the type of secondary school attended); and the child's BMI measured at age 7. The use of such a rich set of control variables minimizes the risk of omitted variable bias when estimating skill production functions.

3. Empirical specification

Our goal is to determine whether there exists a relationship between pubertal development and boys' and girls' math performance. We model the math skill production function using the framework developed by Todd and Wolpin (2007, 2003), which considers school, family, and children's inputs.

Recent literature has demonstrated the existence of 'sensitive periods' in the formation of skills over the life cycle; that is, that human capital inputs are likely to have different effects at different developmental stages (Cunha and Heckman, 2007; Kautz et al., 2014; Keane et al., 2022). There should thus be no *a priori* expectation that the impact of pubertal development on math scores is constant with age. Therefore, as a starting point, we test whether the effect of pubertal development on math skills at 11 years of age is similar to the effect at 16 years, using the following contemporaneous model:

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

where $Test_{it}$ is individual i 's standardized test score at age t ; PU_{it} is a vector of individual i 's level of adolescent development at age t , as measured by a puberty index; B_i is a vector of a wide set of controls, including pregnancy and birth characteristics, family characteristics, school characteristics, and the child's own attributes as described in Table A.2 in Appendix A; and e_{it} is the error term.

As noted by Fiorini and Keane (2014), in estimating Eq. (1) we face the problem of distinguishing between a mere correlation between adolescent development and cognitive achievement, and a true causal effect. A common source of endogeneity is simultaneity (reverse causality). This source of endogeneity is unlikely to be a problem in our case; we do not believe that children will be able to influence their adolescent development based on their scores in cognitive tests. 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 adolescent development and test scores. Given that no dataset contains a complete history of all relevant inputs, omitted variables remain a possibility. 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-effects models and value-added models (see Del Boca et al., 2017; Fiorini and Keane, 2014; Todd and Wolpin, 2007, 2003). Within-child fixed-effects (first-difference) models require that human capital production functions are constant with age, so that both unobserved inputs and the

⁶ For girls, the correlation between the reported age at menarche and the medically assessed puberty index is 0.53 at age 11 and 0.22 at age 16. For boys, the correlation between the reported age at which the voice broke and the puberty index is 0.17 at age 11 and 0.32 at age 16.

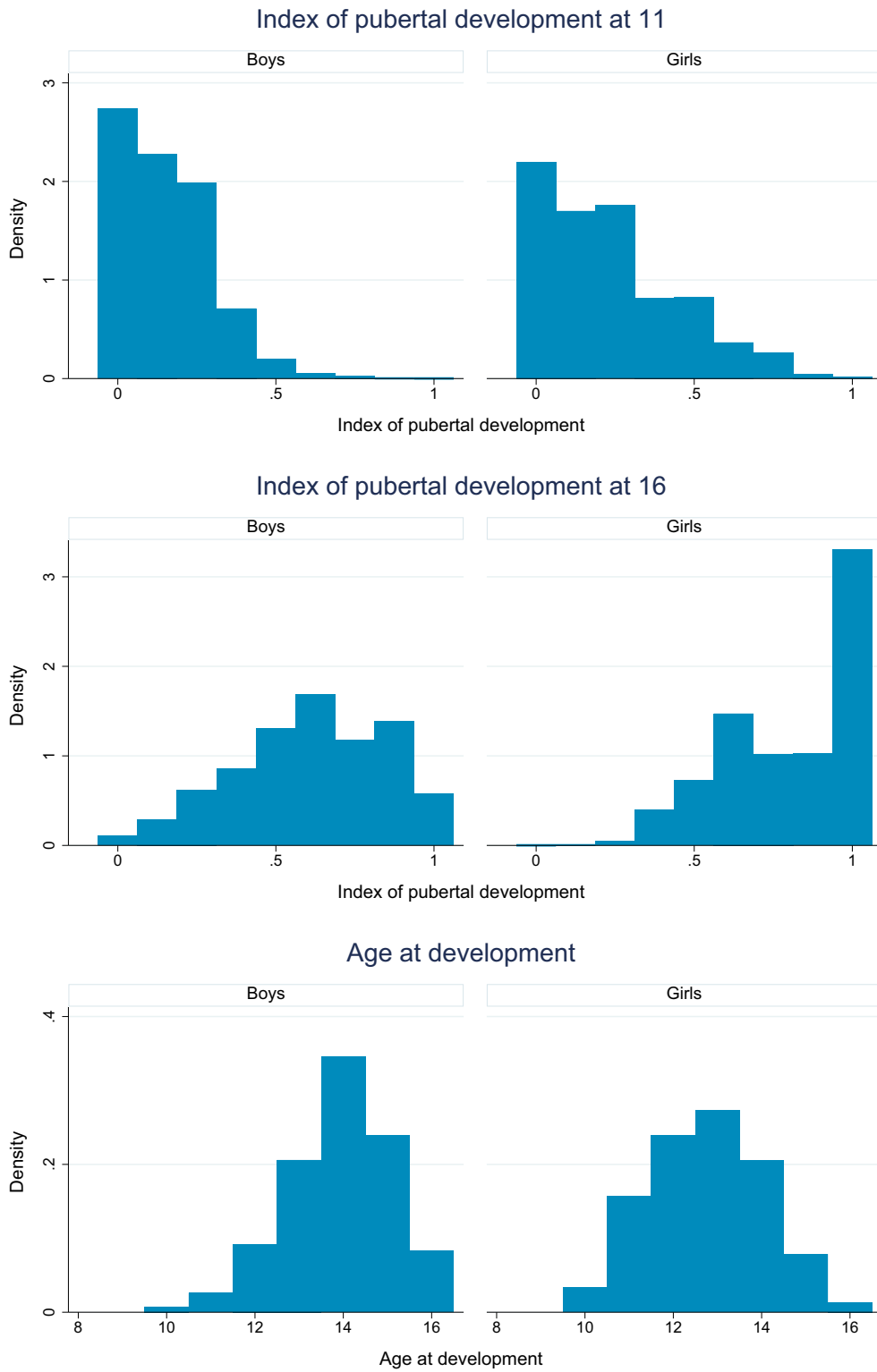


Fig. 2. Pubertal development by gender at ages 11 and 16.
 Source: NCDS, 6461 observations.

Table 2
Impact of pubertal development on math scores. Contemporaneous model.

	(1) Age 11 Boys	(2) Girls	(3) Age 16 Boys	(4) Girls
Puberty Index at t	0.207** (0.082)	0.173*** (0.056)	0.220*** (0.056)	-0.033 (0.063)
No. of observations	5501	5283	3610	3425

Notes: Estimates obtained from the contemporaneous specification in Eq. (1) estimated at ages 11 and 16. Each regression includes basic demographic controls (ethnicity, mother’s age at birth, whether the mother is a lone parent, mother’s education); pregnancy and birth controls (birth order, birth weight, gestation, birth complications, maternal smoking during pregnancy, maternal working before birth); other family controls (English not spoken at home, mother’s height, mother’s BMI); school type; the child’s BMI at age 7; the child’s age in months at assessment; and region of residence at birth. Standard errors in parentheses. Coefficients significant at the * 10%, ** 5% and *** 1% levels. Source: NCDS.

endowment can be differenced out. This assumption is not plausible in the context of sensitive periods during childhood and adolescence. We thus adopt value-added models as our preferred approach, although we also estimate a child fixed-effects model in Section 4.2 as a robustness check.

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 (i) the effects of all inputs, observed and unobserved, and the endowment decline at the same rate with age, and (ii) unobserved current inputs are uncorrelated with observed current inputs. For instance, the first condition requires that the impact of pubertal development at age 11 on test scores at age 16 is a fraction λ_t of the impact of pubertal development at age 11 on test scores at age 11, and the impact of the endowment at age 16 is a fraction λ_t of the impact of the endowment at age 11. If the estimated coefficient on past test scores $\hat{\lambda}_t$ is less than one, the value-added specification implies that the effect of past inputs must always decrease over time. The second condition requires that the set of unobserved current inputs is uncorrelated with the contemporaneous impact of pubertal development at 11 and 16 years of age. We include a wide set of controls B_t to attempt to satisfy this condition.⁷ Formally, our basic VA specification may be written:

$$Test_{it} = PU_{it}\alpha_t + B_t\rho_t + \lambda_t Test_{it-1} + e_{it} \tag{2}$$

We may relax the assumption that the effect of observed inputs on test scores declines at the same rate as the effects of unobserved inputs and the endowment by estimating an extended version of the value-added model that includes lagged inputs. This gives the cumulative value-added (CVA) model (Todd and Wolpin, 2007). Given that we only observe pubertal 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 pubertal development on cognitive skills and allow for different impacts of pubertal development at age 11 on test scores at ages 16 and 11. We do, 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, we specify the CVA model as:

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

Next, we address measurement error in the lagged test score.⁸ In value-added models, measurement error tends to bias downwards the coefficient on lagged achievement (λ_t in Eq. (3)) and may bias the observed input coefficients in an ambiguous direction (Del Bono et al., 2016; Keane et al., 2022). Our preferred specification involves instrumenting for lagged test scores using two-period lagged outcomes in a cumulative value-added instrumental variables (CVA-IV) specification:

$$Test_{it} = \sum_{k=0}^t PU_{it-k}\alpha_k + B_t\rho_t + \lambda_t \widehat{Test}_{it-1} + e_{it} \tag{4}$$

Note that the CVA-IV model can only be estimated for skills at age 16.

4. Results: the dynamic effects of pubertal development and math test scores

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

Table 2 presents estimates of the relationship between the puberty index and math test scores for boys and girls at ages 11 and 16,

⁷ See, however, Section 4.2 below where we address the possibility of potential additional sources of endogeneity.
⁸ We do not address measurement error in pubertal development, first, because doctors’ assessments are relatively unlikely to show measurement errors correlated to test scores; and second, because it is difficult to obtain valid instruments for puberty measures.

Table 3
Impact of pubertal development on math scores. Value-added, cumulative value-added and IV cumulative value-added models.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Age 11		Age 16		Cum. Value-added		Cum. Value-added IV	
	Value-added		Value-added		Boys	Girls	Boys	Girls
	Boys	Girls	Boys	Girls				
Puberty Index at t	0.115 (0.070)	0.095* (0.049)	0.083* (0.043)	-0.095** (0.048)	0.078* (0.046)	-0.117** (0.052)	0.065 (0.048)	-0.140** (0.056)
Puberty Index at $t-1$					-0.027 (0.073)	0.147*** (0.052)	-0.029 (0.077)	0.135** (0.057)
Math scores at $t-1$	0.468*** (0.011)	0.431*** (0.011)	0.674*** (0.012)	0.616*** (0.013)	0.669*** (0.013)	0.620*** (0.014)	0.780*** (0.027)	0.824*** (0.031)
No. of observations	5501	5283	3610	3425	3610	3425	3307	3158

Notes: Estimates obtained from the VA model (age 11) and the VA, CVA and CVA-IV models (age 16).

Each regression includes basic demographic controls (ethnicity, mother's age at birth, whether the mother is a lone parent, mother's education); pregnancy and birth controls (birth order, birth weight, gestation, birth complications, maternal smoking during pregnancy, maternal working before birth); other family controls (English not spoken at home, mother's height, mother's BMI); school type; the child's BMI at age 7; the child's age in months at assessment; and region of residence at birth.

Standard errors in parentheses. Coefficients significant at the * 10%, ** 5% and *** 1% levels.

Source: NCDS.

using the contemporaneous specification in Eq. (1). The estimates show clearly that the association between pubertal development and math test scores differs by gender and age: whereas for boys the association between these variables is very similar at ages 11 and 16, for girls the positive association observed at 11 years is no longer present at 16.

Column 1 shows that a one-point increase in the puberty index at age 11 (the difference between a child who has not started puberty and a completely developed child), is associated with an increase in boys' math scores of 20% of a standard deviation. A more realistic difference in the boys' puberty index of 10 percentage points (eg genital development at stage 2 rather than stage 1, on the 5-stage scale of genital development) would be associated with an increase in math test scores of 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 for girls, a similar increase of 10 percentage points in the puberty index at age 11 (such as breast development at stage 2 rather than stage 1 on a 5-stage scale of breast development) 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. The results at age 16 are shown in Columns 3 and 4. For boys, a 10 percentage point increase in the puberty index (eg pubic hair development at stage 4 instead of stage 2, on the 4-point scale of pubic hair growth) is associated with an increase in math scores for boys of approximately 2% of a standard deviation, but for girls there is no significant effect.

Table 3 presents results from the value-added (VA), cumulative value-added (CVA) and cumulative value-added instrumental-variables (CVA-IV) models described in Eqs. (2), (3) and (4), addressing concerns over potential endogeneity arising from omitted variables and measurement error. The coefficients in Table 3 differ substantially from those in Table 2: the estimated effects are reduced for boys at both ages and for girls at age 11. However, the negative relationship between girls' pubertal development and math scores at age 16, insignificant in Table 2, is significant in columns (4), (6) and (8) of Table 3. The fact that the estimates of the value-added model in Table 3 differ from the estimates of the contemporaneous model in Table 2 suggest that unobserved inputs and the endowment are correlated with pubertal development and the estimates in Table 2 are likely biased.⁹

Columns 1 and 2 of Table 3 show that at 11 years of age, the impact of current pubertal development is no longer significant for boys, and for girls it drops to half the size suggested by the estimates in Table 2: a 10pp increase in the girls' puberty index at age 11 is now associated with an increase in math scores of just 0.9% of a standard deviation. At 16 years of age (Columns 3 and 4), a 10pp increase in the puberty index (such as axillary hair development at stage 4 [adult] instead of stage 3 [intermediate]) 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 fully mature (1) pubertal development at age 16 increases boys' math scores and decreases girls' math scores by roughly 10% of a standard deviation.

Columns 5 and 6 of Table 3 present estimates for the cumulative value-added model of Eq. (3), relaxing the assumption that the effect of pubertal development on math scores declines over time at the same rate as the impacts of unobserved inputs and the

⁹ Table A.3 in Appendix A shows that results remain virtually identical, albeit less precisely estimated, when we use consistent samples throughout and in particular, when the restricted sample of boys and girls for which we have data at age 16 is used to estimate the impact of puberty at age 11.

Table 4
Impact of pubertal development on the math gender gap. Value-added models.

	Age 11 (1) Value-added	Age 16 (2) Value-added	(3) CVA	(4) CVA-IV
Panel A: Without measures of pubertal development				
Female (ref: male)	0.002 (0.014)	-0.179*** (0.014)	-0.179*** (0.014)	-0.171*** (0.015)
Panel B: Including measures of pubertal development				
Female (ref: male)	-0.005 (0.020)	-0.055 (0.048)	-0.081 (0.051)	-0.068 (0.054)
Pubertal development at <i>t</i>	0.104 (0.069)	0.090** (0.042)	0.081* (0.045)	0.06 (0.048)
Pubertal development at <i>t</i> * Girl	-0.011 (0.083)	-0.179*** (0.063)	-0.198*** (0.069)	-0.195*** (0.072)
Pubertal development at <i>t-1</i>			-0.029 (0.072)	-0.03 (0.077)
Pubertal development at <i>t-1</i> * Girl			0.183** (0.088)	0.172* (0.094)
No. of observations	10,793	7727	7727	6465

Notes: Estimates obtained from the VA model (age 11) and the VA, CVA and CVA-IV models (age 16).

Each regression includes basic demographic controls (ethnicity, mother's age at birth, whether the mother is a lone parent, mother's education); pregnancy and birth controls (birth order, birth weight, gestation, birth complications, maternal smoking during pregnancy, maternal working before birth); other family controls (English not spoken at home, mother's height, mother's BMI); school type; the child's BMI at age 7; the child's age in months at assessment; and region of residence at birth.

Standard errors in parentheses. Coefficients significant at the * 10%, ** 5% and *** 1% levels.

Source: NCDS.

endowment. In general, including information on pubertal development at age 11 does not change our estimates of the contemporaneous impact of pubertal development on boys' math scores at age 16, but suggests an even larger negative contemporaneous impact of puberty on girls' math scores at age 16.¹⁰ These estimates suggest that at 16 years of age, a 10pp increase in the puberty index 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, compared with the 0.95% in Column 4. Column 6 additionally shows that early pubertal development has long-term effects beyond its impact through past scores for girls. A 10-percentage-point increase in the puberty index for 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 absent (0) to mature (1) development at age 11.¹¹

Columns 7 and 8 of Table 3 further address measurement problems. Using twice-lagged skills to instrument for lagged test scores increases the coefficient on the lagged test scores as expected by about 20 – 30 percent. The point estimates of the effect of pubertal development on math scores remain very similar, however. As in Columns 3 and 5, pubertal development continues to impact boys' math scores positively, though the effect is no longer significant. For girls, a 10pp increase in the puberty index at age 16 reduces 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.

Overall, the findings in Table 3 indicate that the impact of pubertal development on math outcomes varies by age between boys and girls. Specifically, advanced development at the age of 11 appears to have a beneficial effect on math performance for girls, yet advanced development at 16 years of age appears to impair girls' math performance. By contrast, the math performance of boys generally appears to be positively influenced by development at the age of 16 years, but not necessarily at the age of 11 years. These results suggest that the relationship between the development index and math scores is not linear. This non-linearity could be caused by most boys in the sample being just starting puberty at 11 and most girls having already completed it at 16, leveling the playing field for girls.

To put these results into perspective, Carneiro et al. (2013) report that one additional year of mother's education increases standardized math scores by 3.8% of a standard deviation at ages 12 – 14. We estimate that the difference between absent and mature

¹⁰ Given that the value-added model is nested on the cumulative value-added model, a straightforward test for the validity of the value-added specification is to test whether the coefficients for pubertal development at age 11 are significantly different from zero in the cumulative value-added specification (Todd and Wolpin, 2007). The fact that the contemporaneous impact of pubertal development on boys' math scores at age 16 does not change implies that the value-added model may be adequate to estimate the unbiased impact of pubertal development on boys' math skills. Probably because girls tend to develop earlier than boys, controlling for pubertal development at age 11 reduces the contemporaneous impact of puberty on math scores at age 16 for girls. In this case, the cumulative value-added model would be more appropriate than the value-added model to estimate the unbiased impact of pubertal development on girls' math skills.

¹¹ In our data, the correlation between the puberty indices at ages 11 and 16 is far from perfect. For girls, the correlation between the index of pubertal development at age 11 and the index at age 16 is 0.22. For boys, the correlation between the index of pubertal development at age 11 and the index at age 16 is 0.23. These figures are unlikely to give rise to multicollinearity problems in our estimates.

Table 5
Impact of pubertal development on math scores at age 16. Alternative model specifications and alternative measure of puberty.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	CVA-IV		Two-step CVA		Latent-factor CVA-IV		First-difference	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Panel A: alternative model specifications								
Puberty Index at t	0.065 (0.048)	-0.140** (0.056)	0.117** (0.049)	-0.106* (0.055)	0.072 (0.049)	-0.205*** (0.073)	0.023 (0.049)	-0.134** (0.058)
Puberty Index at $t-1$	-0.029 (0.077)	0.135** (0.057)	-0.019 (0.078)	0.151*** (0.056)	-0.041 (0.079)	0.212*** (0.074)	-0.033 (0.079)	0.133** (0.057)
Math scores at $t-1$	0.780*** (0.027)	0.824*** (0.031)			0.736*** (0.095)	1.233*** (0.139)		
Reading scores at $t-1$					0.064 (0.095)	-0.533*** (0.143)		
Non-cognitive scores at $t-1$					-0.023 (0.027)	-0.002 (0.035)		
Panel B: alternative puberty indicator								
Age at development	0.268* (0.154)	-0.195 (0.133)	0.367** (0.152)	-0.085 (0.132)	0.342** (0.160)	-0.209 (0.170)	0.211 (0.168)	-0.196 (0.138)
Age at development squared	-0.010* (0.006)	0.008 (0.005)	-0.014** (0.006)	0.003 (0.005)	-0.013** (0.006)	0.008 (0.007)	-0.008 (0.006)	0.008 (0.005)
Math scores a $t-1$	0.748*** (0.029)	0.790*** (0.029)			0.586*** (0.101)	1.190*** (0.132)		
Reading scores at $t-1$					0.188* (0.101)	-0.509*** (0.136)		
Non-cognitive scores at $t-1$					-0.006 (0.029)	0.028 (0.034)		
No. of observations	3307	3158	3307	3158	3307	3158	3610	3425

Notes: Estimates obtained from the CVA-IV model (Equation 4); the two-step CVA model (Equation 5); the latent-factor CVA-IV model (Eq. (6)) and the first-difference model (Eq. (7)).

Each regression includes basic demographic controls (ethnicity, mother's age at birth, whether the mother is a lone parent, mother's education); pregnancy and birth controls (birth order, birth weight, gestation, birth complications, maternal smoking during pregnancy, maternal working before birth); other family controls (English not spoken at home, mother's height, mother's BMI); school type; the child's BMI at age 7; the child's age in months at assessment; and region of residence at birth.

Standard errors in parentheses. Coefficients significant at the * 10%, ** 5% and *** 1% levels.

Source: NCDS.

development in girls at age 11 is associated with an increase of about 15% of a standard deviation in their math scores at age 16, and that the difference between absent and mature development in boys at age 16 is associated with an increase in about 7% of a standard deviation in their math scores at the that age. Thus, our estimated effects are comparable in magnitude to the impact of between 1 and 4 additional years of maternal education. The impact of pubertal development on math scores is quantitatively and qualitatively important.

4.1. Puberty and the math gender gap

The results in the previous section estimated the effects of pubertal development on girls' and boys' math scores separately. We now focus on estimating the size of the math gender gap, which we implement by estimating a single equation for girls and boys together, with the inclusion of a gender dummy. Estimating a single equation implicitly assumes that boys and girls share the same production function (see the discussion in Baker and Milligan, 2016).

Table 4 presents estimates obtained from the VA, CVA, and CVA-IV models¹² described in Eqs. (2), (3) and (4). The top panel (Panel A) shows the size of the math gender gap without the inclusion of any puberty measures. The models estimated in Panel B include the puberty index plus its interaction with the gender dummy. Results for age 11 (column 1) show that once previous test scores are controlled for, there is virtually no gender gap in math at 11 years of age, and this finding is unchanged by the inclusion of controls for pubertal development. Results for age 16 (columns 2 – 4) show a gender gap of about 18% of a standard deviation in the model which does not control for pubertal development. This gap diminishes dramatically to almost a third of its previous size and becomes insignificant when pubertal development and its interaction with gender are included in the regressions. In the simplest terms, well over half of the math gender gap at age 16 may be explained by the effects of puberty.

4.2. Robustness checks: alternative model specifications and measures of adolescent development

In this section we assess the sensitivity of our findings to (1) changes in model specification and (2) the use of an alternative self-

¹² We lack information on math performance before age 7 and thus cannot analyse changes in the math gender gap between ages 7 and 11 using any of our value-added models.

assessed measure of pubertal development. We first present two variations on our basic value-added model of the impact of pubertal development at 16 years of age that incorporate additional information on reading scores and non-cognitive skills in estimating math skill production functions, followed by estimates from a child fixed-effects model.

First, [Del Boca et al. \(2017\)](#) propose an additional model that relaxes the requirement that past scores are uncorrelated with the error term in [Eq. \(3\)](#). In their two-step estimation procedure, they first compute an individual fixed-effects estimate of the depreciation rate (λ_t in [Eq. \(3\)](#)), 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 11 and 16 years of age to estimate this persistence parameter that controls for the child-specific endowment. In a second step, λ_t is replaced with its estimate from the fixed-effects regression, obtaining the two-step cumulative value-added model:

$$Test_{it} - \hat{\lambda}_t Test_{it-1} = \sum_{k=0}^t PU_{it-k} \alpha_k + B_i \rho_t + e_{it} \quad (5)$$

Secondly, in the spirit of [Cunha et al. \(2010\)](#), we consider skills as imperfectly measured latent variables and include in the production function for math skills past measures of other skills, namely reading and non-cognitive skills. Following [Agostinelli et al. \(2019\)](#) we address potential mismeasurement of these lagged skill measures by instrumenting them with two-period-lagged skill measures. Thus, our second robustness check involves estimating a variant of [Eq. \(4\)](#) that also incorporates information on past reading and non-cognitive outcomes, adequately instrumented. This latent-factor cumulative value-added IV specification may be written as:

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

Thirdly, following [Todd and Wolpin \(2003\)](#) and [Keane et al. \(2022\)](#) we also estimate a child fixed-effects (first-difference) model. The key assumption behind this model is that the effects of unobserved inputs and endowments are constant over time and can be differenced out. We do, however, allow for the impact of puberty to depend on age as in [Keane et al. \(2022\)](#). This model allows for the effects of puberty to differ by age while controlling for the potential non-randomness of puberty.¹³ Our child fixed-effects model may be written thus:

$$Test_{it} - Test_{it-1} = PU_{it} \alpha_t - PU_{it-1} \alpha_{t-1}^* + v_{it} \quad (7)$$

Panel A in [Table 5](#) presents estimates from the alternate specifications in Equations (5) – (7). For the purposes of comparison, estimates from the CVA-IV model (Columns 7 and 8 of [Table 3](#)) are reproduced in Columns 1 and 2. These estimates remain broadly unchanged under both the two-step and the latent-factor CVA models of [Eqs. \(5\) and \(6\)](#) and the fixed-effects model of [Eq. \(7\)](#).¹⁴ Estimates of the contemporaneous impact of pubertal development on boys' math scores at 16 years of age now range from 0.2% 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 the puberty index at 11 years of age increases girls' math scores by between 1.3% and 2% of a standard deviation.

We then test the sensitivity of our findings to an alternative measure of adolescent development, namely the self-reported age at menarche (girls) and the voice breaking (boys). Panel B of [Table 5](#) considers the association between a quadratic function of age at development, and math scores at age 16, using the CVA-IV model of [Eq. \(4\)](#).¹⁵ Age at development has a convex relationship with girls' math scores and a clearly concave relationship with boys' math scores. The relationship between age at development and boys' math scores presents a maximum between 13.1 and 13.4 years of age, about half a year before the average age at development for 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 would be associated with between 0.3% and 0.7% of a standard deviation higher math score. These results are in line with the effects shown in [Table 3](#) of generally positive effects of puberty on boys' math scores at age 16, and generally negative effects of puberty on math scores for girls.¹⁶

Taken together, the results in this section suggest that the findings reported in [Table 3](#) are quite robust; the estimates obtained from [Eq. \(4\)](#) can be considered a lower bound for the impact of pubertal development on math scores for boys and girls.

¹³ For instance, early life adversity has been shown to be associated with early puberty onset ([Colich et al., 2020](#)). The fixed-effects specification controls for time-invariant sources of pubertal development and ensures that our estimates are not (for example) merely picking up cognitive differences between girls who have, and have not, experienced early adversity. See, however, the discussion in [Section 4.2](#) on the potential time-varying factors affecting the non-randomness of pubertal development.

¹⁴ In the latent-factor cumulative value-added model, we include past reading scores and past non-cognitive scores as additional information to control for unobserved inputs and the endowment. It is hard to interpret these coefficients in any reasonable way due to the high correlation obtained between past math performance and past reading performance: over 0.73 for both boys and girls at 16 years of age.

¹⁵ Most studies using age at development as a measure of pubertal development use non-linear specifications for indicators of early development: for instance [Cavanagh et al. \(2007\)](#) use onset of menarche before age 12. Given the arbitrariness of such thresholds we opted for maintaining non-linearities in a less rigid fashion, using a quadratic function in age at development.

¹⁶ The relationship between age at development and girls' math scores is not significantly different from zero. As previously noted, while voice breaking has been validated as an adequate marker for boys' pubertal development, girls' age at menarche has not.

Table 6
Validation exercise.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Age 11		Age 16					
	OLS		OLS		Cum. OLS		Cum. OLS-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	-0.030	0.030	-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)
No. of observations	5501	5283	3610	3425	3610	3425	3307	3158

Notes: Estimates obtained from regressions analogous to those in Eqs. (2)–(4) with dependent variables as indicated (lagged values of math scores rather than lagged dependent variables are included).

Each regression includes basic demographic controls (ethnicity, mother's age at birth, whether the mother is a lone parent, mother's education); pregnancy and birth controls (birth order, birth weight, gestation, birth complications, maternal smoking during pregnancy, maternal working before birth); other family controls (English not spoken at home, mother's height, mother's BMI); school type; the child's BMI at age 7; the child's age in months at assessment; and region of residence at birth.

Standard errors in parentheses. Coefficients significant at the * 10%, ** 5% and *** 1% levels.

Source: NCDS.

4.3. 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 time-invariant or time-varying sources of endogeneity affecting the impact of pubertal development on math performance. However, it is possible that pubertal development may be related to childhood factors such as childhood trauma or nutrition (Colich et al., 2020; Kaplowitz et al., 2001), and that controlling for past test scores may not be enough to account for these sources of endogeneity as value-added models assume. There is no formal means of testing this assumption. However, we perform a validation exercise by identifying a set of variables which are highly likely to be correlated with any such unobserved factors, and testing whether these variables are associated with pubertal development. We select three such variables. *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). These variables are measured before the start of adolescence, and could not be affected by pubertal development; a significant relationship would indicate that there remains some selection bias not properly accounted for by model specifications.

Results are presented in Table 6, in the form of estimates from OLS models using birth weight, maternal education, and smoking during pregnancy as dependent variables. While pubertal development is generally unrelated to birthweight and smoking during pregnancy, there does seem to be a systematic association between pubertal development at age 16 and maternal education for boys. We test whether this association may be sufficient to affect our estimates of the relationship between pubertal development and math scores. Table A.4 in the Appendix presents results for the same regressions estimated in Table 3, but also controlling for the interaction between maternal education and pubertal development at ages 11 and 16. The fact that the results in Table A.4 are very similar to those presented in Table 3 may be taken as a sign that omitted variable bias is limited (Oster, 2019). Overall, the evidence suggests that selection bias not properly accounted for by model specifications is unlikely to be driving our results.

We continue this line of inquiry by adding into the model a range of covariates that have been shown to be correlated with puberty as well as math outcomes. We test three sets of variables: the first related to family background, the second to school characteristics and the third to physical and psychological characteristics of the child.¹⁷

Family background may impact gender differences in math through at least three potential channels. First, family resources may differently impact boys' and girls' educational outcomes (Autor et al., 2019; Lei and Lundberg, 2020). Second, parents may make differentiated gender-specific investments, particularly time investments (Baker and Milligan, 2016); for example, sons may receive more inputs of time from fathers (Lundberg, 2005). And third, parents' preferences or gender attitudes may be transmitted to their

¹⁷ Table A.5 In Appendix A presents descriptive statistics by gender for the variables used in this section.

children (Dossi et al., 2021). There is evidence on the critical impact of same-sex family figures during adolescence in shaping later roles and choices (Olivetti et al., 2020). The literature shows not only correlations between mothers' and children's attitudes towards working women (Farré and Vella, 2013), but also correlations between mothers' and daughters' labour supply (Olivetti et al., 2020). 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 A.6 compares results from a benchmark regression based on estimates in Table 3, with results from regressions including the three family variables: whether the mother had a job, whether the mother was a manager, and whether the mother expects the child to have a management-level job. Results show that these variables do not affect the estimated impact of pubertal development on boys' and girls' math scores.

Table A.7 presents results from a similar exercise which investigates the role of school environment. School environments may affect the math gender gap, especially via the influence of teachers and peers. Teachers' gender role beliefs and expectations may influence students' self-image, affecting their interests and aspirations and ultimately their academic outcomes, leading to a self-fulfilling prophecy (Carlana, 2019; Papageorge et al., 2020). 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 (Alan et al., 2018; Lavy and Sand, 2018). The gender composition of the classroom has also been studied as a potential explanatory factor in gender differences in math scores. Girls are more likely to conform to gender stereotypes in the presence of boys 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, 2010). The empirical evidence shows that the proportion of boys in the classroom does not affect the gender gap in math, however, except when it considers only the proportion of high-achieving boys (Bharadwaj et al., 2016; Cools et al., 2022). Nonetheless, single-sex schooling has been found to improve girls' academic outcomes in math and math related subjects (Booth et al., 2018; Eisenkopf et al., 2015). The results in Table A.7 show that the inclusion of school inputs (the gender of the teacher; whether the teacher expects the child to follow STEM subjects; the number of math courses taken by the child; and whether the child attends a coeducational school) does not significantly change the estimated impact of puberty on the math scores of either boys or girls.

Table A.8 repeats the same exercise with characteristics of the children themselves. From a biological standpoint, being overweight has been shown to be correlated with academic outcomes (Sabia, 2007). Differences in perseverance, risk aversion, or leadership may also be behind math differences between boys and girls: Ellison & Swanson (2021) show that girls give up more easily; Andreoni et al. (2020) show that they are more risk-averse, and Alan et al. (2020) show that they are less willing to take leadership roles. Table A.8 tests the effects of the child's BMI and a binary indicator of behavioural problems based on the Rutter scale (Rutter, 1967). Results show that neither of these variables alters the estimated relationship between pubertal development and math scores.

In summary, the analysis in Sections 4.2 and 4.3 demonstrates that our estimates of the impact of pubertal development on boys' and girls' math scores in adolescence are robust to a wide range of challenges. We now move on to explore possible mechanisms underlying those estimated relationships.

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

Adolescence may impact gender disparities in mathematical performance through both biological and societal factors. Biological mechanisms could operate via the physical and hormonal changes associated with puberty, if these also affect cognitive abilities and behavior related to math performance (Halpern, 2011).¹⁸ In addition, adolescence is a sensitive period, second only to early childhood, when connections between neurons (synapses) can be pruned or strengthened to increase neural efficiency (Blakemore and Choudhury, 2006; Dahl et al., 2018; Steinberg, 2014).¹⁹ Social mechanisms could operate via increased pressures to conform to stereotypical gender roles, leading to widening gaps in math performance as boys and girls develop differentiated identities and attitudes. Studies have found that this gap can be seen in traits such as competitiveness and risk-taking, with girls becoming less competitive and more risk-averse compared to boys in patriarchal societies (Alan et al., 2020; Andersen et al., 2013; Andreoni et al., 2020; Niederle and Vesterlund, 2010).

The purpose of this paper is not to make a definitive argument as to whether the mathematical performance of boys and girls are determined by nature or nurture, particularly in light of recent discoveries in neuroscience establishing that the social environment plays a crucial role in brain development (Blakemore, 2018). However, it is certainly worth discussing whether our findings point to biological or social mechanisms.

¹⁸ For instance, adolescents' testosterone levels have been shown to relate to spatial ability and competitiveness not only for boys but also for girls (Bateup et al., 2002; Vuoksimaa et al., 2012).

¹⁹ Recent research in neuroscience recognizes that adolescents' brains are physically different from the brains of younger children and adults (Blakemore 2018 p.73). 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 (Blakemore et al., 2010; Giedd et al., 1999). Longitudinal studies suggest that the physical and hormonal changes during puberty may also directly influence the grey matter decreases and white matter increases taking place in adolescence (Goddings et al., 2014; Herting et al., 2015; Piekarski et al., 2017). Brain maturation is accompanied by higher efficiency in cognitive function including improvements in intelligence quotient, working memory, problem solving, executive functioning, and social cognition (Fuhrmann et al., 2015; Steinberg et al., 2009). In consequence pubertal development may directly impact social and cognitive abilities (Picci and Scherf, 2016; Steinberg, 2008), especially executive functioning skills playing a critical role in the development of mathematics proficiency (Cragg and Gilmore, 2014).

Table 7
Comparison of publicly evident vs. non publicly evident markers of pubertal development.

	(1) Age 11 Value-added	(2) Age 16 Value-added	(3) Cum. value-added	(4) Cum. value-added IV
Panel A: Impact of Breast Development Within Pubic Hair Development Categories				
Breast Development Index at t	0.137** (0.062)	-0.112** (0.049)	-0.137*** (0.052)	-0.126** (0.055)
Breast Development Index at $t-1$			0.121** (0.049)	0.097* (0.053)
Panel B: Impact of Pubic Hair Development Within Breast Development Categories				
Pubic Hair Measure at t	-0.037 (0.063)	0.034 (0.057)	0.034 (0.060)	0.007 (0.064)
Pubic Hair Measure at $t-1$			0.135*** (0.046)	0.137*** (0.050)
No. of observations	5283	3425	3425	3158

Notes: Estimates obtained from the VA model (age 11) and the VA, CVA and CVA-IV models (age 16). Panel A includes pubic hair development categories fixed effects. Panel B includes breast development categories fixed effects.

Each regression also includes basic demographic controls (ethnicity, mother's age at birth, whether the mother is a lone parent, mother's education); pregnancy and birth controls (birth order, birth weight, gestation, birth complications, maternal smoking during pregnancy, maternal working before birth); other family controls (English not spoken at home, mother's height, mother's BMI); school type; the child's BMI at age 7; the child's age in months at assessment; and region of residence at birth.

Standard errors in parentheses. Coefficients significant at the * 10%, ** 5% and *** 1% levels.

Source: NCDS.

Our results so far suggest a socially-driven explanation for the widening of the math gender gap in adolescence. If the effect of puberty on math outcomes were due solely to physiological factors, this effect would arguably remain consistent regardless of the child's age. The observed variations in the impact of pubertal development at ages 11 and 16, particularly among girls, suggest a social channel coming into play alongside the physical changes. In what follows we present additional evidence to support the theory that it is socialization and gender roles rather than biological forces that shape how adolescent development impacts the math performance of boys and girls.

We first check whether markers of puberty are differently related to math performance according to whether these markers are generally evident to the public or not. We re-estimate the value-added models in Table 3 using the index of breast development (a publicly evident marker of puberty for girls) and the index of pubic hair development (a marker which is not publicly evident) as explanatory variables.²⁰ Panel A in Table 7 presents estimates of the impact of breast development on math scores within groups of girls with similar stages of pubic hair development; Panel B in Table 7 shows the estimated impact of pubic hair development on math scores within groups of girls at similar stages of breast development.

As shown in Panel A, an increase of 10pp in the breast development index at age 11 is associated with an increase of between 1% and 1.4% of a standard deviation in math scores at ages 11 and 16, while a 10pp increase in the breast development index at age 16 is associated with a reduction of almost 1.4% of a standard deviation in math scores at age 16. Conversely, Panel B shows that we do not find any significant contemporaneous impacts of pubic hair development on girls' math scores at either 11 or 16 years of age. In sum, the results in Table 7 show that our main results in Table 3 are driven by the effects of a publicly evident markers of pubertal development such as girls' breast development, rather than non-publicly evident markers of maturation such as pubic hair growth, pointing towards a social channel for the impact of pubertal development on math outcomes.

Second, we explore heterogeneity in the impacts of pubertal development on math scores by self-assessed math ability. Prior literature emphasizes that self-assessed measures of ability are largely driven by socially gendered stereotypes (Bordalo et al., 2019, 2016; Carlana, 2019; Coffman, 2014). A finding that the impact of pubertal development on math performance differs by self-assessed ability, a socially-driven factor, would be consistent with a social rather than a biological explanation for the impact of puberty on the widening of the gender gap in math. We re-estimate the VA specifications, including a binary indicator of above-average levels of self-assessed math ability, and interactions between this variable and the puberty index. Panel A in Table 8 reveals that there are no significant interactions for girls, but that for boys there is a significant positive interaction between the puberty index at age 16 and above-average self-assessed ability; in other words, the positive impact of pubertal development on boys' math scores is driven by those with higher self-assessed math ability. For boys with above-average self-perceived ability, a 10pp increase in the puberty index at age 16 is associated with an increase in math scores of almost 2% of a standard deviation. The fact that pubertal development positively impacts math scores of boys with above average self-perceived ability, and does not affect the scores of boys with lower self-perceived ability, is again consistent with social factors playing a role in explaining the effect of puberty on math outcomes by gender.

Finally, we explore whether the impact of adolescent development on math scores varies by the level of early skills. A finding of differential impacts of pubertal development on math performance by early skills would be consistent with a greater intrinsic

²⁰ We focus on publicly evident and non-publicly evident measures of development for girls because we lack information on publicly evident measures of development for boys at 11 years of age.

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

	(1) Age 11 Value-added Boys	(2) Age 16 Girls	(3) Cum. value-added Boys	(4) Girls	(5) Cum. value-added IV Boys	(6) Girls
Panel A: By Self-Assessed Math Performance						
Above Average Self-Assessed Performance	0.534*** (0.038)	0.485*** (0.050)	0.394*** (0.074)	0.616*** (0.122)	0.308*** (0.078)	0.423*** (0.126)
Puberty Index at t	0.151** (0.075)	0.072 (0.050)	0.042 (0.049)	-0.119** (0.053)	0.031 (0.052)	-0.151*** (0.057)
Puberty Index at $t-1$			0.041 (0.077)	0.155*** (0.054)	0.028 (0.082)	0.142** (0.059)
Above Average*Puberty Index at t	-0.065 (0.179)	0.137 (0.139)	0.185* (0.109)	-0.019 (0.151)	0.193* (0.112)	0.047 (0.151)
Above Average*Puberty Index at $t-1$		-0.18	-0.1 (0.175)		-0.137 (0.181)	-0.09 (0.142)
Panel B: By Initial Skills						
Top Half at Age 7	0.082* (0.043)	-0.001 (0.044)	0.054 (0.063)	0.127 (0.089)	-0.076 (0.078)	-0.066 (0.099)
Puberty Index at t	0.118 (0.093)	0.149** (0.062)	0.043 (0.056)	-0.138** (0.062)	0.019 (0.063)	-0.181*** (0.069)
Puberty Index at $t-1$			0.058 (0.092)	0.196*** (0.065)	0.081 (0.103)	0.178** (0.076)
Top Half*Puberty Index at t	-0.006 (0.139)	-0.128 (0.094)	0.077 (0.093)	0.044 (0.110)	0.09 (0.097)	0.09 (0.113)
Top Half*Puberty Index at $t-1$			-0.203 (0.148)	-0.105 (0.103)	-0.231 (0.156)	-0.1 (0.110)
No. of observations	5501	5283	3610	3425	3307	3158

Notes: Estimates obtained from the VA model (age 11) and the VA, CVA and CVA-IV models (age 16).

Each regression includes basic demographic controls (ethnicity, mother's age at birth, whether the mother is a lone parent, mother's education); pregnancy and birth controls (birth order, birth weight, gestation, birth complications, maternal smoking during pregnancy, maternal working before birth); other family controls (English not spoken at home, mother's height, mother's BMI); school type; the child's BMI at age 7; the child's age in months at assessment; and region of residence at birth.

Standard errors in parentheses. Coefficients significant at the * 10%, ** 5% and *** 1% levels.

Source: NCDS.

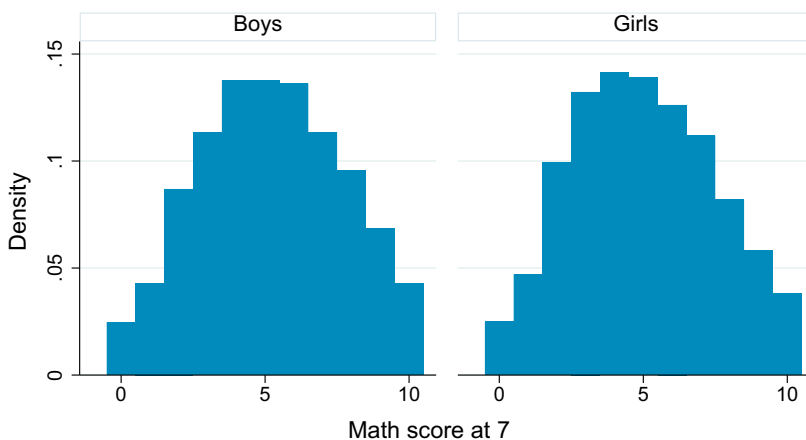
variability in math abilities among males (Ellison and Swanson, 2010; Hyde and Mertz, 2009; Machin and Pekkarinen, 2008).²¹ If we assume that early skills are largely biologically determined, a finding of larger impacts of pubertal development for high-performing boys would suggest a biological channel for the impact of puberty on the gender gap in math. We re-estimate the VA specifications including interactions between the puberty indices and an indicator of having above-average math skills at 7 years of age. The estimates in Panel B of Table 8 show that there are no significant differences in the impact of pubertal development on math scores by this indicator of early skills, showing a lack of support for a biological explanation. Notice, however, that both the point estimates and the confidence intervals for the interaction variables are very large, especially for the impact of pubertal development at 11 years of age on girls' math performance at 11 and boys' math performance at 16. In any case, the evidence found in Panel B of Table 8 regarding differences in the impact of puberty by early skills is not as conclusive as the evidence found in Panel A of Table 8 regarding differences by self-assessed ability, and our results fail to support a clear biological explanation for the impact of pubertal development on the widening of the gender gap in math.

All in all, the evidence we present on the differentiated impacts of pubertal development on math scores by age and gender, together with the fact that more publicly evident measures of pubertal development show larger impacts, and the fact that the impact of pubertal development on boys' math scores is driven by boys with a higher self-perception of their own ability in math, seems to rule out a purely biological explanation for the gender gap in math and suggests that social conditions take precedence over biological factors in shaping the gender gap in math.

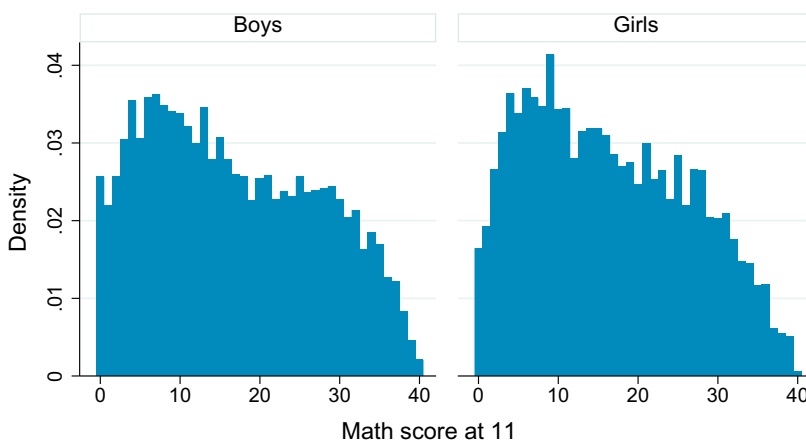
6. Conclusion

This paper has investigated the relationship between adolescent development and the increase in the gender math gap from primary to secondary school. We use data from the NCDS, a unique longitudinal dataset that followed all British children born in a single

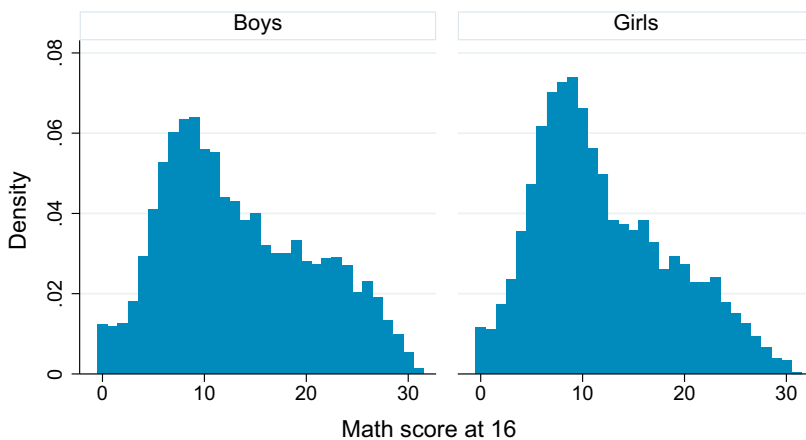
²¹ There is some controversy regarding this greater variability hypothesis. Using PISA data for 2003, Machin and Pekkarinen (2008) find that the average variance ratio (boys' variance over girls' variance) which should be 1.00 if there was equality is 1.13; and is even larger -1.70- when considering the top 5 percent of the distribution. However, Using TIMMs data for 2003 and 2007, Kane and Mertz (2012) present 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' over girls' performance is not found in some countries.



Graphs by female



Graphs by female



Graphs by female

Fig. C.1. Distribution of math test scores by gender at ages 7, 11 and 16. 6461 observations.
 Source: NCDS.

week in 1958, and which offers medical assessments of pubertal development alongside data on children’s cognitive outcomes and a rich set of household, school, and birth characteristics. We document that for this cohort, the gender gap in math widens by about 10% of a standard deviation over the adolescent years. Using dynamic production function models, we show that pubertal development can

explain almost two thirds of the increased gap in math scores between boys and girls. Our results are robust to the use of a range of 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 math gender gap. The fact that impacts are larger for publicly evident measures of pubertal development, and that they vary with the ages of boys and girls and with the degree of self-perceived math ability, allow us to rule out a purely biological channel for this effect. Consistent with recent neurobiological theories that emphasize that brain synapses are pruned in response to environmental influences (Dahl et al., 2018), our results imply a sociologically driven explanation for the gender gap in math.

Our finding that the widening of the math gender gap during adolescence is partly due to socially-driven factors associated with the timing of puberty has important policy implications. It suggests, for instance, that interventions aimed at lessening gender stereotyping (for example, informational or promotional campaigns to highlight the importance of math for later professional development, or positive role models for girls in math-related fields) have the potential to be effective in addressing the math gap; it suggests also that the timing of interventions to support math achievement in girls matters, and that such interventions may need to be targeted earlier rather than later, and certainly before puberty is complete.

Our study is based on data from the National Child Development Study (NCDS) cohort, who were aged 16 in 1974. Given that puberty is starting earlier in the lifecycle (Farello et al., 2019), it is possible that the gender gap in math may also start to widen at earlier ages in more recent cohorts. This has important implications for educators and policymakers, as it suggests that interventions to support academic achievement may need to be implemented at even earlier ages than are suggested by the results in this paper. Further research is needed to determine the generalizability of our findings to more contemporaneous cohorts and to explore the relationship between puberty and academic achievement in more depth.

Declaration of Competing Interest

The authors declare no conflict of interest.

Data availability

We have shared the link to the data [link](#).

Acknowledgements

Borra acknowledges funding from Grant RTI2018-098217-B-I00 funded by MCIN/AEI/ 10.13039/501100011033 and by “ERDF A way of making Europe”. Sevilla acknowledges support from the European Research Council (CoG PARENTTIME-ERC-2017). This research is based on data from the 1958 National Child Development Study (NCDS). We are grateful to the Centre for Longitudinal Studies (CLS), UCL Social Research Institute, for the use of these data and to the UK Data Service for making them available. However, neither CLS nor the UK Data Service bear any responsibility for the analysis or interpretation of these data.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.euroecorev.2023.104542](https://doi.org/10.1016/j.euroecorev.2023.104542).

Appendix B. Examples of questions in math tests administered to NCDS cohort

Tests administered at age 7

- Peter had 4 toy cars and he 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?

Tests administered at age 11

- In a class of 40 pupils $\frac{3}{4}$ 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.

Tests administered at age 16

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

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

- b. $x = 7y = 3$
- c. $x = 5y = 3$
- d. $x = 2y = -2$
- e. $x = 6y = 2$

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

- a. n^2
- b. $2n$
- c. $2n - 1$
- d. $2n + 1$
- e. $n^2 - 2$

Appendix C. Distribution of scores in NCDS math tests

References

- Agostinelli, F., Saharkhiz, M., Wiswall, M., 2019. Home and School in the Development of Children (No. w26037). National Bureau of Economic Research, Cambridge, MA. <https://doi.org/10.3386/w26037>.
- Aizer, A., 2017. A review essay on Isabel Sawhill's generation unbound: drifting into sex and parenting without marriage and Laurence Steinberg's age of opportunity: lessons from the new science of adolescence. *J. Econ. Lit.* 55, 592–608. <https://doi.org/10.1257/jel.20151337>.
- Alan, S., Ertac, S., Kubilay, E., Loranth, G., 2020. Understanding gender differences in leadership. *Econ. J.* 130, 263–289. <https://doi.org/10.1093/ej/uez050>.
- Alan, S., Ertac, S., Mumcu, I., 2018. Gender stereotypes in the classroom and effects on achievement. *Rev. Econ. Stat.* 100, 876–890. https://doi.org/10.1162/rest_a.00756.
- Andersen, S., Ertaç, S., Gneezy, U., List, J., Maximiano, S., 2013. Gender, competitiveness, and socialization at a young age: evidence from a matrilineal and a patriarchal society. *Rev. Econ. Stat.* 95, 1438–1443. https://doi.org/10.1162/REST_a.00312.
- Andrabi, T., Das, J., Khwaja, A.I., Zajonc, T., 2011. Do value-added estimates add value? Accounting for learning dynamics. *Am. Econ. J.* 3, 29–54. <https://doi.org/10.1257/app.3.3.29>.
- Andreoni, J., Di Girolamo, A., List, J.A., Mackevicius, C., Samek, A., 2020. Risk preferences of children and adolescents in relation to gender, cognitive skills, soft skills, and executive functions. *J. Econ. Behav. Organ.* 179, 729–742. <https://doi.org/10.1016/j.jebo.2019.05.002>.
- Autor, D., Figlio, D., Karbownik, K., Roth, J., Wasserman, M., 2019. Family disadvantage and the gender gap in behavioral and educational outcomes. *Am. Econ. J.* 11, 338–381. <https://doi.org/10.1257/app.20170571>.
- Baker, M., Milligan, K., 2016. Boy-girl differences in parental time investments: evidence from three countries. *J. Hum. Cap.* 10, 399–441. <https://doi.org/10.3386/w18893>.
- Bateup, H.S., Booth, A., Shirtcliff, E.A., Granger, D.A., 2002. Testosterone, cortisol, and women's competition. *Evol. Hum. Behav.* 23, 181–192. [https://doi.org/10.1016/S1090-5138\(01\)00100-3](https://doi.org/10.1016/S1090-5138(01)00100-3).
- Bharadwaj, P., De Giorgi, G., Hansen, D., Neilson, C., 2016. The gender gap in mathematics: evidence from Chile. *Econ. Dev. Cult. Change* 65, 141–166. <https://doi.org/10.1086/687983>.
- Black, D., Haviland, A., Sanders, S.G., Taylor, L., 2008. Gender wage disparities among the highly educated. *J. Hum. Resour.* 43, 630–659. <https://doi.org/10.3368/jhr.43.3.630>.
- Blakemore, S., Burnett, S., Dahl, R.E., 2010. The role of puberty in the developing adolescent brain. *Hum. Brain Mapp.* 31, 926–933. <https://doi.org/10.1002/hbm.21052>.
- Blakemore, S.J., 2018. *Inventing Ourselves: the Secret Life of the Teenage Brain*. PublicAffairs, New York.
- Blakemore, S.J., Choudhury, S., 2006. Development of the adolescent brain: implications for executive function and social cognition. *J. Child Psychol. Psychiatry* 47, 296–312. <https://doi.org/10.1111/j.1469-7610.2006.01611.x>.
- Blau, F.D., Kahn, L.M., 2017. The gender wage gap: extent, trends, and explanations. *J. Econ. Lit.* 55, 789–865. <https://doi.org/10.1257/jel.20160995>.
- Booth, A., Cardona-Sosa, L., Nolen, P., 2018. Do single-sex classes affect academic achievement? An experiment in a coeducational university. *J. Public Econ.* 168, 109–126. <https://doi.org/10.1016/j.jpubeco.2018.08.016>.
- Bordalo, P., Coffman, K., Gennaioli, N., Shleifer, A., 2016. Stereotypes. *Q. J. Econ.* 131, 1753–1794. <https://doi.org/10.1093/qje/qjw029>.
- Bordalo, P., Coffman, K., Gennaioli, N., Shleifer, A., 2019. Beliefs about gender. *Am. Econ. Rev.* 109, 739–773. <https://doi.org/10.1257/aer.20170007>.
- Borgonovi, F., Choi, A., Paccagnella, M., 2021. The evolution of gender gaps in numeracy and literacy between childhood and young adulthood. *Econ. Educ. Rev.* 82, 102119. <https://doi.org/10.1016/j.econedurev.2021.102119>.
- Brooks-Gunn, J., Warren, M.P., Rosso, J., Gargiulo, J., 1987. Validity of self-report measures of girls' pubertal status. *Child Dev.* 58, 829–841. <https://doi.org/10.2307/1130220>.
- Card, D., Payne, A.A., 2021. High school choices and the gender gap in stem. *Econ. Inq.* 59, 9–28. <https://doi.org/10.1111/ecin.12934>.
- Carlana, M., 2019. Implicit stereotypes: evidence from teachers' gender bias. *Q. J. Econ.* 134, 1163–1224. <https://doi.org/10.1093/qje/qjz008>.
- Carneiro, P., Meghir, C., Pary, M., 2013. Maternal education, home environments, and the development of children and adolescents. *J. Eur. Econ. Assoc.* 11, 123–160. <https://doi.org/10.1111/j.1542-4774.2012.01096.x>.
- Case, A., Paxson, C., 2008. Stature and status: height, ability, and labor market outcomes. *J. Polit. Econ.* 116, 499–532. <https://doi.org/10.1086/589524>.
- Cavanagh, S.E., Riegle-Crumb, C., Crosnoe, R., 2007. Puberty and the education of girls. *Soc. Psychol. Q.* 70, 186–198. <https://doi.org/10.1177/019027250707000207>.
- Ceci, S.J., Ginther, D.K., Kahn, S., Williams, W.M., 2014. Women in academic science: a changing landscape. *Psychol. Sci. Public Interest* 15, 75–141. <https://doi.org/10.1177/1529100614541236>.
- Centre for Longitudinal Studies, 2023. National Child Development Study: Childhood Data from Birth to Age 16, Sweeps 0-3, 1958-1974, 3rd Edition. SN: 5565. <http://doi.org/10.5255/UKDA-SN-5565-2>.
- Coffman, K., 2014. Evidence on self-stereotyping and the contribution of ideas. *Q. J. Econ.* 129, 1625–1660. <https://doi.org/10.1093/qje/qju023>.
- Coleman, L., Coleman, J., 2002. The measurement of puberty: a review. *J. Adolesc.* 25, 535–550. <https://doi.org/10.1006/jado.2002.0494>.
- Colich, N.L., Platt, J.M., Keyes, K.M., Sumner, J.A., Allen, N.B., McLaughlin, K.A., 2020. Earlier age at menarche as a transdiagnostic mechanism linking childhood trauma with multiple forms of psychopathology in adolescent girls. *Psychol. Med.* 50, 1090–1098. <https://doi.org/10.1017/S0033291719000953>.

- Contini, D., Tommaso, M.L.D., Mendolia, S., 2017. The gender gap in mathematics achievement: evidence from Italian data. *Econ. Educ. Rev.* 58, 32–42. <https://doi.org/10.1016/j.econeduc.2017.03.001>.
- Cools, A., Fernández, R., Patacchini, E., 2022. The asymmetric gender effects of high flyers. *Labour Econ.* 79 <https://doi.org/10.1016/j.labeco.2022.102287>.
- Cragg, L., Gilmore, C., 2014. Skills underlying mathematics: the role of executive function in the development of mathematics proficiency. *Trends Neurosci. Educ.* 3, 63–68. <https://doi.org/10.1016/j.tine.2013.12.001>.
- Cunha, F., Heckman, J., 2007. The technology of skill formation. *Am. Econ. Rev.* 97, 31–47. <https://doi.org/10.1257/aer.97.2.31>.
- Cunha, F., Heckman, J.J., Schennach, S.M., 2010. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica* 78, 883–931. <https://doi.org/10.3982/ECTA6551>.
- Dahl, R.E., Allen, N.B., Wilbrecht, L., Suleiman, A.B., 2018. Importance of investing in adolescence from a developmental science perspective. *Nature* 554, 441–450. <https://doi.org/10.1038/nature25770>.
- Day, F.R., Elks, C.E., Murray, A., Ong, K.K., Perry, J.R.B., 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. <https://doi.org/10.1038/srep11208>.
- Del Boca, D., Flinn, C., Verriest, E., Wiswall, M., 2019. Actors in the child development process. [10.3386/w25596](https://doi.org/10.3386/w25596).
- Del Boca, D., Monfardini, C., Nicoletti, C., 2017. Parental and child time investments and the cognitive development of adolescents. *J. Labor Econ.* 35, 565–608. <https://doi.org/10.1086/689479>.
- Del Bono, E., Francesconi, M., Kelly, Y., Sacker, A., 2016. Early maternal time investment and early child outcomes. *Econ. J.* 126, F96–F135. <https://doi.org/10.1111/ecej.12342>.
- Doepke, M., Sorrenti, G., Zilibotti, F., 2019. The economics of parenting. *Annu. Rev. Econom.* 11, 55–84. <https://doi.org/10.1146/annurev-economics-080218-030156>.
- Dossi, G., Figlio, D., Giuliano, P., Sapienza, P., 2021. Born in the family: preferences for boys and the gender gap in math. *J. Econ. Behav. Organ.* 183, 175–188. <https://doi.org/10.1016/j.jebo.2020.12.012>.
- Eisenkopf, G., Hessami, Z., Fischbacher, U., Ursprung, H., 2015. Academic performance and single-sex schooling: evidence from a natural experiment in Switzerland. *J. Econ. Behav. Organ.* 115, 123–143. <https://doi.org/10.1016/j.jebo.2014.08.004>.
- Ellison, G., Swanson, A., 2010. The gender gap in secondary school mathematics at high achievement levels: evidence from the American Mathematics Competitions. *J. Econ. Perspect.* 24, 109–128. <https://doi.org/10.1257/jep.24.2.109>.
- Ellison, G., Swanson, A., 2021. Dynamics of the gender gap in high math achievement. *J. Human Resour.* <https://doi.org/10.3368/jhr.58.5.0620-10972R1>.
- Farello, G., Altieri, C., Cutini, M., Pozzobon, G., Verrotti, A., 2019. Review of the literature on current changes in the timing of pubertal development and the incomplete forms of early puberty. *Front. Pediatr.* 7, 147 <https://doi.org/10.3389/fped.2019.00147>.
- Farré, L., Vella, F., 2013. The intergenerational transmission of gender role attitudes and its implications for female labour force participation. *Economica* 80, 219–247. <https://doi.org/10.1111/ecca.12008>.
- Figlio, D., Guryan, J., Karbownik, K., Roth, J., 2014. The effects of poor neonatal health on children's cognitive development. *Am. Econ. Rev.* 104, 3921–3955. <https://doi.org/10.1257/aer.104.12.3921>.
- Fiorini, M., Keane, M., 2014. How the allocation of children's time affects cognitive and noncognitive development. *J. Labor Econ.* 32, 787–836. <https://doi.org/10.1086/677232>.
- Fryer, R.G., Levitt, S.D., 2010. An empirical analysis of the gender gap in mathematics. *Am. Econ. J.* 2, 210–240. <https://doi.org/10.1257/app.2.2.210>.
- Fuhrmann, D., Knoll, L.J., Blakemore, S.J., 2015. Adolescence as a sensitive period of brain development. *Trends Cogn. Sci.* 19, 558–566. <https://doi.org/10.1016/j.tics.2015.07.008>.
- 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. <https://doi.org/10.1038/13158>.
- Gneezy, U., Niederle, M., Rustichini, A., 2003. Performance in competitive environments: gender differences*. *Q. J. Econ.* 118, 1049–1074. <https://doi.org/10.1162/00335530306098496>.
- 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. <https://doi.org/10.1016/j.neuroimage.2013.09.073>.
- Goldin, C., Katz, L.F., Kuziemko, I., 2006. The Homecoming of American college women: the reversal of the college gender gap. *J. Econ. Perspect.* 20, 133–156. <https://doi.org/10.1257/jep.20.4.133>.
- Goodman, J., 2014. The wages of sinistrality: handedness, brain structure, and human capital accumulation. *J. Econ. Perspect.* 28, 193–212. <https://doi.org/10.1257/jep.28.4.193>.
- Greiner, D.J., Rubin, D.B., 2011. Causal effects of perceived immutable characteristics. *Rev. Econ. Stat.* 93, 775–785. <https://www.jstor.org/stable/23016076>.
- Guiso, L., Monte, F., Sapienza, P., Zingales, L., 2008. Culture, gender, and math. *Science* 320, 1164–1165. <https://doi.org/10.1126/science.1154094>.
- Guryan, J., Ludwig, J., Bhatt, M.P., Cook, P.J., Davis, J.M.V., Dodge, K., Farkas, G., Fryer Jr., R.G., Mayer, S., Pollack, H., Steinberg, L., Stoddard, G., 2023. Not too late: improving academic outcomes among adolescents. *Am. Econ. Rev.* 113, 738–765. <https://doi.org/10.1257/aer.20210434>.
- Halpern, D.F., 2011. Sex Differences in Cognitive Abilities, 4th Edition. Psychology Press, New York. <https://doi.org/10.4324/9780203816530>.
- Heckman, J., Pinto, R., 2015. Causal analysis after Haavelmo. *Econ. Theory* 31, 115–151. <https://doi.org/10.1017/S026646661400022X>.
- Herting, M.M., Paltam, P., Spielberg, J.M., Dahl, R.E., Sowell, E.R., 2015. A longitudinal study: changes in cortical thickness and surface area during pubertal maturation. *PLoS ONE* 10, e0119774. <https://doi.org/10.1371/journal.pone.0119774>.
- Hodges-Simeon, C.R., Gurven, M., Cárdenas, R.A., Gaulin, S.J.C., 2013. Voice change as a new measure of male pubertal timing: a study among Bolivian adolescents. *Ann. Hum. Biol.* 40, 209–219. <https://doi.org/10.3109/03014460.2012.759622>.
- Holland, P.W., 1986. Statistics and causal inference. *J. Am. Stat. Assoc.* 81, 945–960. <https://doi.org/10.2307/2289064>.
- Hyde, J.S., Mertz, J.E., 2009. Gender, culture, and mathematics performance. *Proc. Natl. Acad. Sci. U. S. A.* 106, 8801–8807. <https://doi.org/10.1073/pnas.0901265106>.
- Iriberrí, N., Rey-Biel, P., 2019. Competitive pressure widens the gender gap in performance: evidence from a two-stage competition in mathematics. *Econ. J.* 129, 1863–1893. <https://doi.org/10.1111/ecej.12617>.
- Joensen, J.S., Nielsen, H.S., 2016. Mathematics and gender: heterogeneity in causes and consequences. *Econ. J.* 126, 1129–1163. <https://doi.org/10.1111/ecej.12191>.
- Kahn, S., Ginther, D., 2018. Women and science, technology, engineering, and mathematics (STEM): are differences in education and careers due to stereotypes, interests, or family? In: Averett, S.L., Argys, L.M., Hoffman, S.D. (Eds.), *The Oxford Handbook of Women and the Economy*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190628963.013.13>, p. 0.
- Kane, J., Mertz, J., 2012. Debunking myths about gender and mathematics performance. *No. Am. Math. Soc.* 59, 10–21. <https://doi.org/10.1090/noti790>.
- Kaplowitz, P.B., Slora, E.J., Wasserman, R.C., Pedlow, S.E., Herman-Giddens, M.E., 2001. Earlier onset of puberty in girls: relation to increased body mass index and race. *Pediatrics* 108, 347–353. <https://doi.org/10.1542/peds.108.2.347>.
- Kautz, T., Heckman, J.J., Diris, R., ter Weel, B., Borghans, L., 2014. Fostering and measuring skills: improving cognitive and non-cognitive skills to promote lifetime success. *OECD Educ. Work. Pap. Series*. <https://doi.org/10.3386/w20749>.
- Keane, M., Krutikova, S., Neal, T., 2022. Child work and cognitive development: results from four low to middle income countries. *Quant. Econom.* 13, 425–465. <https://doi.org/10.3982/QE1745>.
- Koerselman, K., Pekkarinen, T., 2018. Cognitive consequences of the timing of puberty. *Labour Econ.* 54, 1–13. <https://doi.org/10.1016/j.labeco.2018.05.001>.
- Lavy, V., Sand, E., 2018. On the origins of gender gaps in human capital: short- and long-term consequences of teachers' biases. *J. Public Econ.* 167, 263–279. <https://doi.org/10.1016/j.jpubeco.2018.09.007>.
- Lei, Z., Lundberg, S., 2020. Vulnerable boys: short-term and long-term gender differences in the impacts of adolescent disadvantage. *J. Econ. Behav. Organ.* 178, 424–448. <https://doi.org/10.1016/j.jebo.2020.07.020>.

- Lundberg, S., 2005. Sons, daughters, and parental behaviour. *Oxford Rev. Econ. Policy* 21, 340–356. <https://doi.org/10.1093/oxrep/gri020>.
- Machin, S., McNally, S., 2005. Gender and student achievement in English schools. *Oxford Rev. Econ. Policy* 21, 357–372. <https://www.jstor.org/stable/23606826>.
- Machin, S., Pekkarinen, T., 2008. Global sex differences in test score variability. *Science* 322, 1331–1332. <https://doi.org/10.1126/science.1162573>.
- Maloney, E.A., Ramirez, G., Gunderson, E.A., Levine, S.C., Beilock, S.L., 2015. Intergenerational effects of parents' math anxiety on children's math achievement and anxiety. *Psychol. Sci.* 26, 1480–1488. <https://doi.org/10.1177/0956797615592630>.
- Martin, A.J., Steinbeck, K., 2017. The role of puberty in students' academic motivation and achievement. *Learn. Individ. Differ.* 53, 37–46. <https://doi.org/10.1016/j.lindif.2016.11.003>.
- Mullis, I.V.S., Martin, M.O., Foy, P., Hooper, M., 2016. *TIMSS 2015 International Results in Mathematics*. Boston College.
- Niederle, M., Vesterlund, L., 2010. Explaining the gender gap in math test scores: the role of competition. *J. Econ. Perspect.* 24, 129–144. <https://doi.org/10.1257/jep.24.2.129>.
- OECD, 2010. PISA 2009 results: what students know and can do: student performance in reading. *Math. Sci. I* <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. [10.1787/9789264201118-en](https://doi.org/10.1787/9789264201118-en).
- OECD, 2019a. Adult education level. [10.1787/36bce3fe-en](https://doi.org/10.1787/36bce3fe-en).
- OECD, 2019b. PISA 2018 Results (Volume II): Where All Students Can Succeed. [10.1787/b5fd1b8f-en](https://doi.org/10.1787/b5fd1b8f-en).
- Olivetti, C., Patacchini, E., Zenou, Y., 2020. Mothers, peers, and gender-role identity. *J. Eur. Econ. Assoc.* 18, 266–301. <https://doi.org/10.1093/jeea/jvy050>.
- Oster, E., 2019. Unobservable selection and coefficient stability: theory and evidence. *J. Bus. Econ. Stat.* 37, 187–204. <https://doi.org/10.1080/07350015.2016.1227711>.
- Papageorge, N., Gershenson, S., Kang, K.M., 2020. Teacher expectations matter. *Rev. Econ. Stat.* 102, 234–251. https://doi.org/10.1162/rest_a_00838.
- Persico, N., Postlewaite, A., Silverman, D., 2004. The effect of adolescent experience on labor market outcomes: the case of height. *J. Polit. Econ.* 112, 1019–1053. <https://doi.org/10.1086/422566>.
- Picci, G., Scherf, K.S., 2016. From caregivers to peers: puberty shapes human face perception. *Psychol. Sci.* 27, 1461–1473. <https://doi.org/10.1177/0956797616663142>.
- Piekarski, D.J., Johnson, C.M., Boivin, J.R., Thomas, A.W., Lin, W.C., Delevich, K., M Galarce, E., Wilbrecht, L., 2017. Does puberty mark a transition in sensitive periods for plasticity in the associative neocortex? *Brain Res.* 1654, 123–144. <https://doi.org/10.1016/j.brainres.2016.08.042>.
- Rodríguez-Planas, N., Sanz-de-Galdeano, A., Terskaya, A., 2022. Gender norms in high school: impacts on risky behaviors from adolescence to adulthood. *J. Econ. Behav. Organ.* 196, 429–456. <https://doi.org/10.1016/j.jebo.2022.01.015>.
- Rutter, M., 1967. A children's behaviour questionnaire for completion by teachers: preliminary findings. *Child Psychol. Psychiatry Allied Discip.* 8, 1–11. <https://doi.org/10.1111/j.1469-7610.1967.tb02175.x>.
- Sabia, J., 2007. The effect of body weight on adolescent academic performance. *South Econ. J.* 73, 871–900. <https://doi.org/10.2307/20111933>.
- Schurer, S., 2017. Does education strengthen the life skills of adolescents? *IZA World Labor*. <https://doi.org/10.15185/izawol.366>.
- Steele, C.M., 1997. A threat in the air. How stereotypes shape intellectual identity and performance. *Am. Psychol.* 52, 613–629. <https://doi.org/10.1037//0003-066x.52.6.613>.
- Steinberg, L., 2008. A social neuroscience perspective on adolescent risk-taking. *Dev. Rev.* 28, 78–106. <https://doi.org/10.1016/j.dr.2007.08.002>.
- Steinberg, L., 2014. *Age of Opportunity: Lessons from the New Science of Adolescence*, *Age of Opportunity: Lessons from the New Science of Adolescence*. Houghton Mifflin Harcourt, Boston, MA.
- 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, 28–44. <https://doi.org/10.1111/j.1467-8624.2008.01244.x>.
- Todd, P.E., Wolpin, K.I., 2003. On the specification and estimation of the production function for cognitive achievement. *Econ. J.* 113, F3–F33. <https://doi.org/10.1111/1468-0297.00097>.
- Todd, P.E., Wolpin, K.I., 2007. The production of cognitive achievement in children: home, school, and racial test score gaps. *J. Hum. Cap.* 1, 91–136. <https://doi.org/10.1086/526401>.
- Vuoksimaa, E., Kaprio, J., Eriksson, C.J.P., Rose, R.J., 2012. Pubertal testosterone predicts mental rotation performance of young adult males. *Psychoneuroendocrinology* 37, 1791–1800. <https://doi.org/10.1016/j.psyneuen.2012.03.013>.
- Walker, I.V., Smith, C.R., Davies, J.H., Inskip, H.M., Baird, J., 2020. Methods for determining pubertal status in research studies: literature review and opinions of experts and adolescents. *J. Dev. Orig. Health Dis.* 11, 168–187. <https://doi.org/10.1017/S2040174419000254>.
- Wehby, G.L., Prater, K., McCarthy, A.M., Castilla, E.E., Murray, J.C., 2011. The impact of maternal smoking during pregnancy on early child neurodevelopment. *J. Hum. Cap.* 5, 207–254. <https://doi.org/10.1086/660885>.