Evidence on the interaction between gender, adolescence, and comprehensive schooling

Kristian Koerselman and Tuomas Pekkarinen

May 16, 2014

Abstract

We study the association between adolescent development and gender differences in educational attainment in selective and comprehensive school systems. We argue that the comprehensive system, in which students are selected into academic and vocational tracks relatively late, is harmful for a subset of boys who lag behind in their adolescent development. We use data from British National Child Development Study, which contains individual-level information on the timing of puberty, educational outcomes, and the age of track selection. We find that late pubertal development seems harmful specifically for boys in late selecting, comprehensive schools, and conclude that the interaction between pubertal development and comprehensive schooling is likely to reinforce the already existent female dominance in education.
Keywords: *XXXX*.
JEL: *XXXX*

1 Introduction

Around the world, women outperform men in terms of educational outcomes. Gender differences in self-discipline and other noncognitive skills have been shown to be able to explain a large part of this outcome gap. (Jacob, 2002; Silverman, 2003; Duckworth & Seligman, 2006)

We also know that there is a strong relationship between the age at which students are first selected into vocational and academic tracks, and the proportion of girls in those academic tracks (Figure 1). At least part of this relationship is likely to be causal, and there is causal evidence that the

Finnish comprehensive school reform, which delayed selection from age 10–11 to 15–16, exacerbated gender differences in the likelihood of completing the academic track in high school. (Pekkarinen, 2008)

In this paper, we argue that these two phenomena may be related. A lack of self-discipline may not be equally harmful at every point in the student's educational career. School systems typically require additional effort and self-discipline at specific, high-stakes points (cf. Jürges et al., 2012; Haraldsvik, 2012; Koerselman, 2013). Track selection typically constitutes the first high-stakes moment in students' educational careers.

During adolescence, a number of changes occur in cognitive and noncognitive functioning, among them a permanent downward change in discount rates (Steinberg et al., 2009). Gender differences in the timing of cognitive and noncognitive development may interact with the timing of high-stakes points. If for example track selection occurs at a point when most girls have made this transition, but some boys have not, this will tend to reinforce existing gender differences in educational outcomes.

To our knowledge, the link between gender differences in the timing of the development of cognitive and noncognitive skills, late selection, and educational outcomes has not been studied before. We therefore turn to a British data set that contains information on the timing of pubertal development of individuals who attended secondary school during a time at which Britain was in the process of moving from a selective system to a comprehensive one.

Our results show that the association between late pubertal development and age 16 test scores or age 23 attainment is is substantially and significantly more negative for boys than for girls, even after adjusting for age 7 test scores and father's social status. Additionally, we show that almost all of this difference is concentrated in the comprehensive schools in our sample. In selective schools, the late development penalty is about equally small for boys as for girls.

We interpret these result as suggestive evidence that late selecting, comprehensive school systems work to reinforce the gender differences already present in educational outcomes by locking in the negative effects of delayed cognitive and noncognitive development.

2 Background

Puberty refers to the physical changes that occur during the transition between childhood and adulthood. It is well known that the timing of the onset

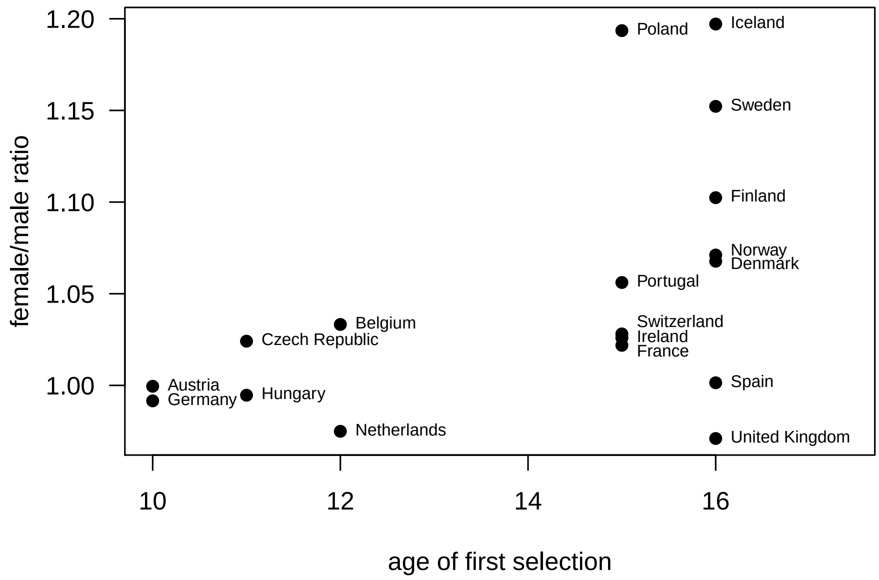


Figure 1: Female/male ratio in academic track and tracking age in Europe. Reproduced from Pekkarinen (2012).

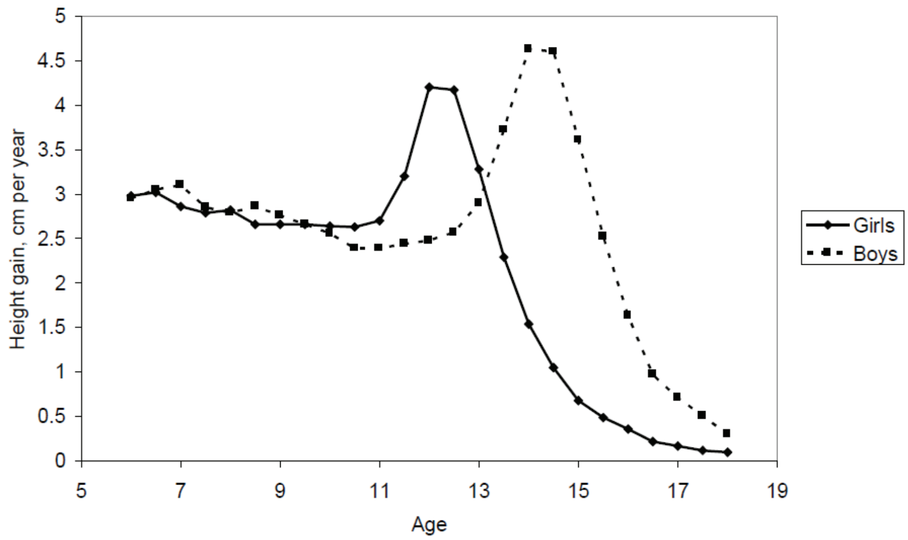


Figure 2: Adolescent spurt in height growth for normally maturing boys and girls. Figure 2 from Tanner (1962) data from Shuttelworth (1939)

of puberty differs by gender, with boys trailing girls by about two years. At age 15–16, most girls have emerged from puberty, while a sizable portion of boys are typically still are undergoing important physical changes. (Tanner, 1962; Steinberg, 2014)

The timing of puberty is affected by environmental variables such as nutrition and health (Steinberg, 2014), and the direct effects of environmental variables on education as well as other, correlated variables such as parental background therefore have the potential to induce a spurious correlation between pubertal development and educational outcomes.

Evidence on the direct effect of pubertal development educational and other outcomes is elusive. Though puberty is regulated by a hormonal feedback loop, the direct effects of hormones on behavior seem limited, especially in mid to late puberty. Causal effects are however hard to identify, not least because of the multitude of potential interactions and lags. (cf. Buchanan et al., 1992)

Even the mere idea that puberty coincides with a period of great turmoil, stress and behavioral problems, is only partially supported by empirical evidence, and it is easy to overstate how common such problems are. (cf. Arnett, 1999; Steinberg & Morris, 2001; Steinberg, 2014)

Adolescence generally refers to an age span that begins with the onset of puberty and ends in early adulthood. During adolescence, important changes occur in cognitive and noncognitive skills, among others in deductive reasoning and metacognition. (XXX) Furthermore, discount rates fall rapidly between ages 12–13 and 16–17, and adolescents become more future-oriented during these ages (Steinberg et al., 2009), perhaps in part because adolescents get better att thinking about counterfactual outcomes,

The relationship between the timing of puberty and the timing of changes in cognitive and noncognitive functions is under-researched. Studies on the timing of the physical changes associated with puberty are usually carried out with time measured since the start of puberty, while studies on cognitive and noncognitive development measure time since birth.

Though evidence is limited, the timing of brain development does seem to be related to the timing of puberty (Blakemore et al., 2010) rather than to age since birth.

Our data set contains variables measuring the timing of puberty, but these variables are best seen as imperfect proxies of associated changes in cognitive and noncognitive functions.

3 Data and methods

The survey most appropriate for our purposes is the longitudinal National Child Development Study (University of London, 2008) or NCDS. It aims to follow all those born in Great Britain in the week starting on the 3rd of March 1958. Follow up survey sweeps have been carried out in 1965, 1969, 1974, 1981, 1991, 1999, 2004 and 2008.

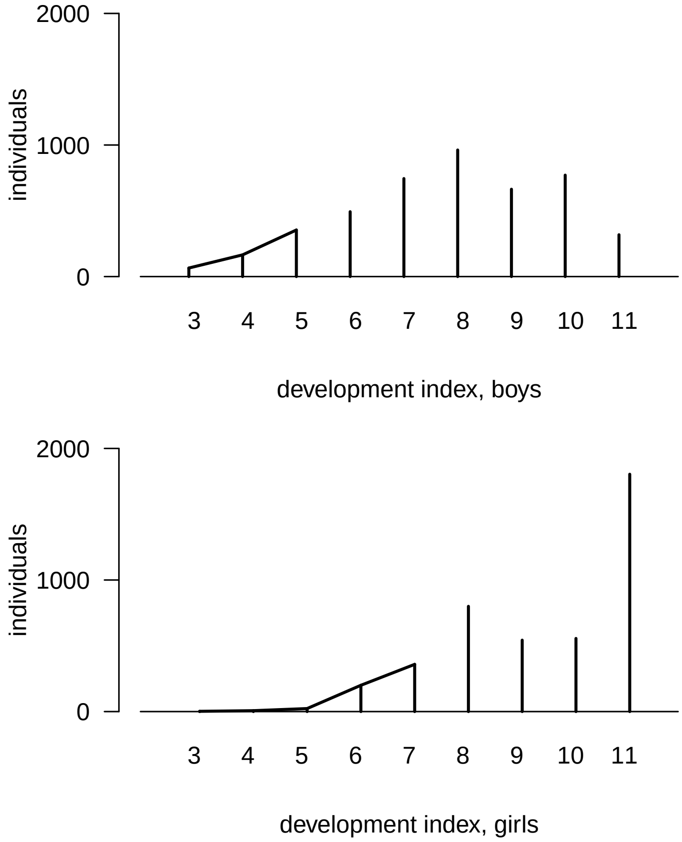
In the 1974 sweep, at age 16, a medical officer collected information on health, among others on the development of pubic and axillary hair for both genders (1 absent; 2 sparse; 3 intermediate; 4 adult), on facial hair for boys (1 absent; 2 sparse; 3 adult) and on breast development for girls (1 absent; 2 intermediate; 3 adult). These follow Tanner's stages of pubertal development (Tanner, 1962).

When interpreted as cardinal, the pubertal development measures are strongly correlated, and a principal component analysis reveals almost identical factor loadings for all three measures for both genders. For the sake of transparency, we therefore simply use the sum of the three measures to create a pubertal development index without losing much in terms of efficiency compared to using the first principal component of the subscores.

The distribution of our pubertal development index can be seen from Figure 3. Boys (top) have much lower values of the index than girls. To avoid potential problems with inter-gender comparisons of pubertal development, we define a student as late developing in relative terms, and run our regressions interacted with gender. For for boys, we use a cutoff of 5 or lower, for girls with a score of 7 or lower. One reason for selecting these cutoffs is that they divide both genders into similarly-sized groups: about 14% of boys and about 13% of girls are below the cutoff. The cutoffs have been illustrated in the figure as well.

As documented in Koerselman (2013), the NCDS cohort was exposed to different tracking policies depending on where they lived. Students attending the tracked system (Figure 4, top half) took the co-called *Eleven Plus* achievement test around age 11. Those who did well enough on the Eleven Plus were allowed to enter a upper track *grammar school*. Grammar school students could acquire an Ordinary Level General Certificate of Education or O-level at age 16, and an Advanced Level General Certificate of Education or A-level at age 18, after which they could enter higher education. Students who failed to qualify for grammar school usually entered a vocational *secondary modern*. Secondary modern students could either acquire a Certificate of Secondary Education or CSE at age 16, or leave the education system

Figure 3: Histograms of the age 16 pubertal development index.



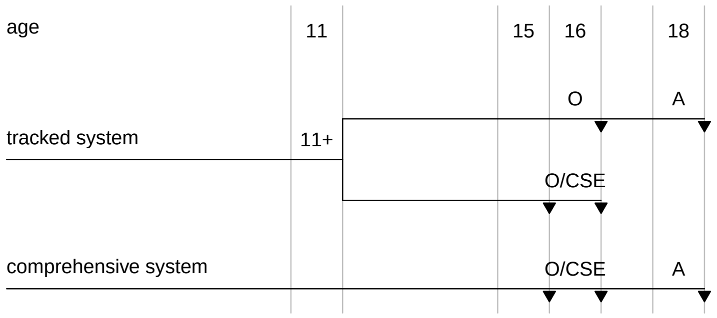
Shown are separate histogrames for boys (top) and for girls. More than 40% of girls have the highest possible score on the index, while boys score much lower. We define boys with an index of 5 or lower, and girls with an index of 7 or lower as late developing.

one year earlier. The highest grade on the CSE was regarded as equivalent to an O-level.

For students in the comprehensive system, selection did not take place at age 11, but rather through voluntary exit at the compulsory schooling age of 15, or otherwise at the CSE or O-level examination at age 16. We have illustrated this in the bottom half of Figure 4.

At age 16, the sample students were administered a mathematics and a reading test, and we use the resulting test scores as outcome measures. The reading test was designed to test for particularly poor reading skills, and the reading scores are strongly negatively skewed as a consequence. To avoid overweigting poor readers, and to aid in the interpretation of regression re-

Figure 4: The main British secondary school systems around 1969.



In the tracked system (top), an age 11 test determined access to the upper track. In the comprehensive system (bottom), all students attended the same middle school. Triangles indicate common exits from secondary education: at the end of compulsory education but before taking the O-level or CSE examinations at the end of middle school; after completing middle school; after taking the A-level examinations at the end of high school. Figure reproduced from Koerselman (2013)

sult, we transform these test scores to standard normal distributions.

We also add information on educational attainment at ages 23 and 46 from the fourth and seventh sweep of the NCDS respecively. For the seventh sweep this information is cumulative, so that the information would for example also be available for an individual missing from the seventh sweep, but not from the sixth. As a consequence, information on attainment at age 46 is available for more individuals than at age 23.

We use the achievement of at least two A-levels in school as an indication of high school completion at age 23. For age 46 we also use an indicator of whether the invidivual has completed a nonvocational higher education degree corresponding to ISCED level 3 for high school completion, and 5 or higher for university.

We use z-scores of age 7 tests and teacher ratings as control variables in some of the specificatons. These include the results of a word recognition and word comprehension test, a copying designs test to assess perceptuo-motor abilities, a draw-a-man test to assess general mental and perceptual abilities, and an arithmetic test. We also control for father's social status.

As can be seen from Table 1, attrition is quite high in the NCDS. When we require information to be known for gender, age 7 test scores, age 11 geographical information as well as for age 16 measures of pubertal development, less then half of the original student sample remains, even if all but one LEA are still represented in the data. The effective sample size drops further when we only compare students who spent all of their middle school years in a tracked school with those spending all of their middle school years in a comprehensive one. We thus drop students that experienced a regime change during their school career, or that attended other types of schools. The latter caterory includes non-government schools.

Table 1: NCDS sample sizes

	students	LEAs
original data	18558	180
key age 7, 11 and 16 information available	8831	179
...of which tracked	2649	128
...of which comprehensive	1960	130
...of which reformed during middle school	2573	130
...of which other school type	1649	161
of students in tracked or comprehensive schools:		
...age 16 scores known	4425	166
...educational attainment age 23 known	2548	163
...educational attainment age 46 known	3018	164

Information from sweeps at ages 23 and 46 is even more sparse. In order to not reduce the sample size too much, we do not require students to have information on all five outcome variables in order to be included in our regressions. This implies that estimates for the different outcomes are not strictly comparable to each other because they are not estimated off the exact same student samples.

We run regression models that are variations of

$$\begin{aligned} y_i &= \alpha + \beta_1 I(\text{male})_i + \beta_2 I(\text{late})_i + \beta_3 I(\text{compr.})_i \\ &\quad + \beta_4 I(\text{male})_i I(\text{late})_i + \beta_5 I(\text{male})_i I(\text{compr.})_i + \beta_6 I(\text{late})_i I(\text{compr.})_i \\ &\quad + \beta_7 I(\text{male})_i I(\text{late})_i I(\text{compr.})_i + \beta_8 X_i + \varepsilon_i \end{aligned}$$

where y_i is an educational outcome, $I()_i$ is an indicator of the student being male, being a late developer and attending a comprehinsive school respectively, and X_i is an (optional) matrix of student background controls. The main coefficient of interest is the one on the triple interaction, β_7, since it gives the conditional mean outcome difference specifically for late developing boys in comprehensive schools. This is the coefficient we hypothetise is negative.

Because student outcomes may be correlated within LEAs, we weight each individual observation with the inverse of the square root of the number of observations within the LEA. This has the same effect as weighting each LEA with the square root of the number of observations contained in it, and should improve efficiency. We also cluster standard errors on the LEA level.

4 Results

We first run a simple regression interacting gender with late development. Results can be seen from Table 2. Late developing girls are estimated to perform either slightly better or slightly worse than their peers, depending on which outcome we look at. For boys however, the relationship is much more dramatic. Compared to girls, the penalty for developing late in terms of age 16 test scores is about 0.2 standard deviations larger. The penalty in terms of the probability of attaining a high school degree by age 23 is almost 10 percentage points larger. All these differences are substantial in size, and significantly different from zero. Differences in age 46 attainment are smaller, but of the same sign.

Table 2: Regression estimates for the unconditional relationship between gender, late pubertal development, and educational outcomes.

	outcome				
	ma16	re16	hs23	hs46	un46
male	0.243	0.114	0.061	0.028	0.035
	(0.027)	(0.028)	(0.016)	(0.016)	(0.014)
late developing	-0.027	-0.087	0.036	-0.029	0.000
	(0.056)	(0.053)	(0.030)	(0.028)	(0.025)
male × late developing	-0.173	-0.203	-0.097	-0.052	-0.050
	(0.076)	(0.069)	(0.044)	(0.040)	(0.037)
controls	no	no	no	no	no
no. students	4425	4442	2548	3018	3018
no. LEAs	166	166	163	164	164

Notes: Outcomes are age 16 math scores, age 16 reading scores, high school at age 23, high school at age 46, and university at age 46. Individual observations are weigthed by the inverse of the square root of within-LEA sample size. Standard errors are clustered at the LEA-level, and have been added in parentheses.

If we add age 7 test scores as well as indicatiors of father's social status to the regression, the coefficient on the difference between the penalty for late developing boys and girls becomes smaller. This can be seen from Table 3. The differences are however still sizable, and for the three first outcomes, are still significantly different from zero.

Table 3: The relationship between gender, late pubertal development, and educational outcomes.

	outcome				
	ma16	re16	hs23	hs46	un46
male	0.247	0.139	0.048	0.025	0.031
	(0.023)	(0.022)	(0.015)	(0.015)	(0.014)
late developing	0.034	-0.023	0.037	-0.014	0.013
	(0.044)	(0.044)	(0.027)	(0.025)	(0.022)
male × late developing	-0.125	-0.157	-0.088	-0.040	-0.042
	(0.061)	(0.057)	(0.041)	(0.036)	(0.033)
controls	yes	yes	yes	yes	yes
no. students	4425	4442	2548	3018	3018
no. LEAs	166	166	163	164	164

Notes: Estimates adjusted for age 7 test scores and father's social status. See Figure 2 for details.

Tables 4 and 5 show estimates from the full model. Gender and late development are now additionally interacted with tracking status. As expected, the penalty for late development is the highest for boys in comprehensive schools. This is true whether we control for ability and father's status (Table 5), or not (Table 4).

Even if the individual interactions are often not significantly different from zero in the full model, we find a remarkibly consistent pattern if we predict group averages from the estimates. In Figure 5, we have illustrated late development penalties for the different outcome variables for boys (black bars) and girls (grey bars) from Table 5. Note that unlike in the regression tables, in this figure the attainment outcomes are expressed in sample standard deviations as well to make them more comparable to the test scores.

As can be seen from the figure, the estimates of the penalty are small and on both sides of zero for girls in either school type as well as for boys in tracked schools. The only group for which penalties are large and consisitent are boys in comprehensive schools. In other words: the significant differences between the late development penalty for boys and girls that we found in Tables 2 and 3 are almost entirely due to boys attending comprehensive schools.

Table 4: The unconditional relationship between gender, late pubertal development, tracking, and educational outcomes.

	outcome				
	ma16	re16	hs23	hs46	un46
male	0.232	0.095	0.051	0.033	0.024
	(0.053)	(0.049)	(0.032)	(0.031)	(0.028)
late developing	0.001	0.006	-0.023	-0.039	-0.026
	(0.101)	(0.098)	(0.047)	(0.054)	(0.044)
comprehensive	-0.198	-0.268	-0.036	-0.005	-0.032
	(0.056)	(0.056)	(0.031)	(0.030)	(0.026)
male × late developing	-0.049	-0.181	0.025	0.033	0.068
	(0.136)	(0.126)	(0.079)	(0.077)	(0.067)
male × comprehensive	-0.055	0.020	-0.033	-0.067	-0.050
	(0.078)	(0.072)	(0.046)	(0.048)	(0.043)
late developing × comprehensive	0.086	-0.031	0.122	-0.013	-0.007
	(0.151)	(0.127)	(0.078)	(0.080)	(0.069)
male × late developing × comprehensive	-0.249	-0.185	-0.200	-0.071	-0.097
	(0.208)	(0.168)	(0.116)	(0.109)	(0.096)
controls	no	no	no	no	no
no. students	4425	4442	2548	3018	3018
no. LEAs	166	166	163	164	164

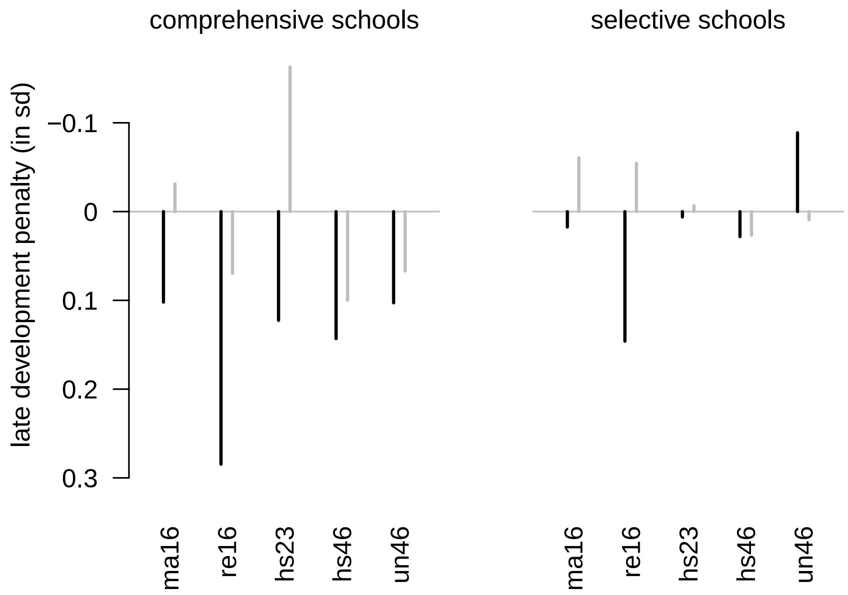
Notes: See Figure 2 for details.

Table 5: The conditional relationship between gender, late pubertal development, tracking, and educational outcomes.

	outcome				
	ma16	re16	hs23	hs46	un46
male	0.276	0.154	0.063	0.045	0.034
	(0.042)	(0.038)	(0.029)	(0.028)	(0.026)
late developing	0.061	0.054	0.003	-0.013	-0.004
	(0.078)	(0.082)	(0.038)	(0.047)	(0.040)
comprehensive	-0.064	-0.152	0.027	0.041	0.001
	(0.046)	(0.048)	(0.029)	(0.026)	(0.024)
male × late developing	-0.078	-0.200	-0.006	-0.001	0.042
	(0.114)	(0.099)	(0.072)	(0.073)	(0.064)
male × comprehensive	-0.127	-0.036	-0.076	-0.083	-0.060
	(0.065)	(0.060)	(0.044)	(0.043)	(0.040)
late developing × comprehensive	-0.029	-0.124	0.070	-0.035	-0.025
	(0.118)	(0.104)	(0.068)	(0.069)	(0.062)
male × late developing × comprehensive	-0.055	-0.015	-0.123	-0.020	-0.057
	(0.169)	(0.144)	(0.109)	(0.101)	(0.090)
controls	yes	yes	yes	yes	yes
no. students	4425	4442	2548	3018	3018
no. LEAs	166	166	163	164	164

Notes: Estimates adjusted for age 7 test scores and father's social status. See Figure 2 for details.

Figure 5: Regression-adjusted differences in educational outcomes for different subgroups.



Regression-adjusted differences in age 16 math and reading scores, highschool completion at ages 26 and 46, as well as university completion at age 46. Shown is the outcome penalty for late developing boys (in black) and girls (in grey) by school system. All outcomes expressed in sample standard deviations. The association between late development and educational outcomes is particularly strong among boys in comprehensive schools.

5 Discussion

Previous research by Pekkarinen (2008, 2012) suggests a causal relationship between comprehensive schooling and the gender gap in educational attainment.

In the current paper, we add a dimension to these findings by interacting gender and comprehensive schooling with information on the timing of individuals' pubertal development. There are theoretical reasons to think that comperhensive school systems should harm late developing boys, who may be locked into educational careers at an age at which they are lagging behind in their cognitive and noncognitive development.

Our analysis is made possible by the richness of the data available in the British National Child Development Study. Unfortunately however, the sample size of the NCDS is not large enough to measure effects of LEA-level institutional factors on educational outcomes with any kind of precision. This is regrettable, but we are not aware of any other data set that combines measures of pubertal development with variation in tracking policies and information on educational outcomes.

We find that late developing boys do significantly worse than late developing girls, and furthermore that this effect is concentrated in the comprehensive schools in our sample. This suggests that the interaction between the timing of adolescence and the timing of track selection may at the very least be a partial explanation of why girls tend to do better in comprehensive school systems.

References

Arnett, J. J. (1999). Adolescent storm and stress, reconsidered. *American psychologist*, *54*, 317.

Blakemore, S.-J., Burnett, S., & Dahl, R. E. (2010). The role of puberty in the developing adolescent brain. *Human brain mapping*, *31*, 926–933.

Buchanan, C. M., Eccles, J. S., & Becker, J. B. (1992). Are adolescents the victims of raging hormones? evidence for activational effects of hormones on moods and behavior at adolescence. *Psychological bulletin*, *111*, 62.

Duckworth, A. L., & Seligman, M. E. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of educational psychology*, *98*, 198.

Haraldsvik, M. (2012). *Influences on Educational Outcomes: Three Essays on the Role of Parents, Peers and Choice*. Ph.D. thesis Norwegian University of Science and Technology.

Jacob, B. (2002). Where the boys aren't: non-cognitive skills, returns to school and the gender gap in higher education. *Economics of Education review*, *21*, 589–598.

Jürges, H., Schneider, K., Senkbeil, M., & Carstensen, C. H. (2012). Assessment drives learning: The effect of central exit exams on curricular knowledge and mathematical literacy. *Economics of Education Review*, *31*, 56–65.

Koerselman, K. (2013). Incentives from curriculum tracking. *Economics of Education Review*, *32*, 140–150.

Pekkarinen, T. (2008). Gender differences in educational attainment: Evidence on the role of tracking from a finnish quasi-experiment. *The Scandinavian Journal of Economics*, *110*, 807–825.

Pekkarinen, T. (2012). Gender differences in education. *Nordic Economic Policy Review*, *2*, 165–196.

Silverman, I. W. (2003). Gender differences in delay of gratification: A meta-analysis. *Sex roles*, *49*, 451–463.

Steinberg, L. (2014). *Adolescence*. McGraw-Hill.

Steinberg, L., Graham, S., O'Brien, L., Woolard, J., Cauffman, E., & Banich, M. (2009). Age differences in future orientation and delay discounting. *Child development*, *80*, 28–44.

Steinberg, L., & Morris, A. (2001). Adolescent development. *Journal of Cognitive Education and Psychology*, *2*, 55–87.

Tanner, J. (1962). *Growth at adolescence.*. Springfield, Ill.

University of London. Institute of Education. Centre for Longitudinal Studies (2008). National Child Development Study: Local Authority Data, 1958-1974: Special Licence Access [computer file]. 2nd Edition. Colchester, Essex: UK Data Archive [distributor], August 2008. SN: 5744.