

IZA DP No. 8693

**School Inputs and Skills:
Complementarity and Self-Productivity**

Cheti Nicoletti
Birgitta Rabe

December 2014

School Inputs and Skills: Complementarity and Self-Productivity

Cheti Nicoletti

*DERS, University of York,
ISER, University of Essex and IZA*

Birgitta Rabe

ISER, University of Essex

Discussion Paper No. 8693
December 2014

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

School Inputs and Skills: Complementarity and Self-Productivity

Using administrative data on schools in England, we estimate an education production model of cognitive skills at the end of secondary school. We provide empirical evidence of self-productivity of skills and of complementarity between secondary school inputs and skills at the end of primary school. Our inference relies on idiosyncratic variation in school expenditure and child fixed effect estimation that controls for the endogeneity of past skills. The persistence in cognitive ability is 0.221 and the return to school expenditure is three times higher for students at the top of the past attainment distribution than for those at the bottom.

JEL Classification: I22, I24

Keywords: education production function, test scores, school quality, complementarity

Corresponding author:

Cheti Nicoletti
Department of Economics and Related Studies
University of York
Heslington, York YO10 5DD
United Kingdom
E-mail: cheti.nicoletti@york.ac.uk

NON-TECHNICAL SUMMARY

This paper looks at whether school expenditure in secondary schools in England is more productive for students with high test scores at the end of primary school than for students with low test scores. Moreover, it examines to what extent skills attained at the end of primary school persist into secondary school (“self-productivity”). These questions are highly relevant for any policy that allocates funds to schools with the aim of reducing inequalities in society. If inputs into highly skilled students are most productive and self-productivity of skills is high this means that policy makers must choose between the most effective use of funds and equality considerations. It also indicates that funds should be allocated in favour of the early school years, because skills attained at one stage have multiplier effects into the future, so that any inequalities between pupils will exacerbate over time.

Despite considerable interest in the effectiveness of school spending, previous research has not been able to answer these questions. This is because of methodological issues in uncovering the causal effects of school spending on progress for children with different levels of skills. We are able to address these issues. First, we can compare the outcomes of similar schools with different funding levels because the way school funding was allocated to schools in England in the time-period covered in our study (2007-2010) resulted in similar schools receiving sometimes very different amounts of funding. Second, we use estimation techniques that allow us to net out unwanted confounders from our effects of interest.

We find that the persistence of cognitive ability at the end of primary school to the end of secondary schooling is 22%. We also find that school inputs are considerably more productive for the brightest students at the end of primary school: the return to school investments is three times higher for these students than for those who were in the bottom 10% at the end of primary school. The implication for policies that allocate funding to schools is to adjust the balance between primary and secondary school in favour of the earlier years.

1 Introduction

In this paper we provide evidence of complementarity between school inputs and skills by estimating an education production model and allowing the return to school investments at a specific stage in the child's life to depend on the level of skills observed at a previous stage. The presence and strength of complementarity has important implications for any policy that allocates funds to schools with the aim of reducing inequalities in society, such as No Child Left Behind in the U.S. If inputs into highly skilled students are most productive, as the child development literature predicts, policy makers face an equity-efficiency trade-off which will exacerbate over the school years through the processes of self-productivity and complementarity. This would indicate that funds should be allocated in favor of the early school years. However, a recent survey of empirical work on the effect of school inputs concludes that increases in resources are often found to be more effective in disadvantaged schools and/or on disadvantaged students at all phases of schooling (Gibbons and McNally, 2013), indicating that remediation is possible and equity and efficiency considerations do not collide in schools.

Despite considerable advances in the estimation of models of skill formation (see Cunha and Heckman 2008; Cunha et al. 2010), solid empirical evidence that could shed light on this is difficult to come by. The estimation framework introduced by Cunha and Heckman (2008) and Cunha et al. (2010) provides clear methods to make inference on complementarities and self-productivity of child skills and can be extended to consider school inputs as well as family investments (see for details Cunha 2011); but, because of scarcity of data providing adequate information on both school and family investments, empirical studies have generally focused mainly either on school inputs or family investments. Furthermore, as noted by Todd and Wolpin (2003), the literature on children's cognitive achievements seems to be split in two separate strands: the early child development strand, which focuses mainly on the effect of family investments, and the educational production model strand, which is more interested in the return to school inputs.

Our paper belongs to the educational production model strand of the literature, indeed our main focus is on evaluating the effect of school inputs on child cognitive abilities using school administrative data that provide details on school inputs and child test scores but

without information on family investments. However, contrary to previous papers in this literature, we aim at making inference on complementarities between school inputs and past cognitive skills. Furthermore, we take account of criticisms raised by the recent literature on child skill formation by addressing the issue of endogeneity of both school inputs and past cognitive skills that could bias inference on dynamic complementarity.¹ As emphasized by Almond and Mazumder (2013), causal inference on dynamic complementarity constitutes a double challenge because it requires both exogenous variation in past abilities and exogenous variation in child investments. This is a situation not often encountered in observational studies. By exploiting idiosyncratic variation in expenditure and controlling for the endogeneity of past ability by using child fixed effect estimation, this paper is able to meet the double challenge of identification and to provide, for the first time, empirical evidence on the causal effect of school inputs in presence of complementarity.

We use rich administrative data on English state schools to evaluate the effect of expenditure per student on the production of cognitive skills of children at the end of compulsory schooling. We consider a value added model that allows students' cognitive skills at the end of compulsory schooling to depend on their cognitive abilities observed at the end of primary school (see for example Hanushek 1986; Hanushek et al. 1996; Todd and Wolpin 2003). Furthermore, we let the effect of expenditure per student vary across children with different levels of past cognitive abilities and we attribute this differential effect to complementarity. We account for possible sorting into schools with different funding levels by using a sample of siblings that go to the same school. We find robust evidence of both self-productivity of cognitive ability and of complementarity. The persistence in cognitive ability is 0.221 and the return to school expenditure is 9% of a standard deviation for students at the top of the past attainment distribution and 3% for those at the bottom.

In order to identify complementarity we need exogenous variation in school inputs. Identification of the effect of school inputs is complicated by the fact that allocation of funding to schools is generally not random and usually designed to decrease social inequalities by favoring children from disadvantaged backgrounds. We are able to take account of the redistributive allocation of school resources by controlling for the school characteristics that

¹Two further issues that have been emphasized in the recent literature of child skill formation are the measurement error in investments and skills and the absence of a natural metric for the test scores measuring child skills (e.g. Cunha 2011). We address these two issues in our section on robustness checks.

determine the allocation of funds to schools. After controlling for these characteristics and a time trend in expenditure, there is idiosyncratic variation in expenditure within and across schools caused by an anomaly in funding rules in England whereby funding is partly detached from educational need. More details on this are given in Section 2.3.2.

In order to identify complementarity we also need exogenous variation in past abilities. Past cognitive abilities are endogenous in the education production model because both current and past cognitive abilities may depend on unobserved cognitive child endowments, unobserved family inputs and other unobserved child characteristics (see Todd and Wolpin 2003, 2007 and Andrabi et al. 2011 for a discussion of the endogeneity issue in dynamic child development models). We are able to control for the potential influence of these unobserved variables by adopting child fixed effect estimation, which is similar in spirit to the within-pupil between-subject estimation used by Dee (2005) and (2007), Clotfelter et al. (2010) and Slater et al. (2010). More precisely, by exploiting the multiplicity of test scores available in different subjects for the same student, we can provide evidence on complementarity by testing whether the return to school inputs increases with the level of lagged cognitive ability.

The child fixed effect estimation allows us to address another criticism raised against some studies on complementarity. The criticism is that skills are multidimensional, and failure to account for this may falsely attribute higher returns to school inputs for children with higher levels of lagged cognitive ability to complementarity (Almond and Mazumder 2013). This is because omitted child capabilities, such as non-cognitive abilities and health, are likely correlated with past and current cognitive abilities. There is ample empirical evidence showing that child abilities are indeed multidimensional and that factors such as non-cognitive abilities are important determinants of children's outcomes (see Heckman et al. 2006; Borghans et al. 2008). Child fixed effect estimation controls for subject-invariant differences in unobserved child skills, be they cognitive, non-cognitive or other skills. It also eliminates other confounding influences, such as peer spillover effects, that are not subject specific. By purging our estimates of subject-invariant endogeneity we are going further than most of the previous literature. However, we cannot rule out that there may be residual subject-specific endogeneity affecting our results. We present a series of checks on our data that indicate that our inference on the presence of self-productivity and dynamic complementarity is not invalidated by subject-specific endogeneity.

The importance of self-productivity and complementarity in the production of skills has been laid out by economists in the child development literature (see Cunha and Heckman 2007, 2008; Cunha et al. 2006, 2010; Aizer and Cunha 2012). There is empirical evidence on complementarity of family investments over the life cycle. For example, several authors have looked at the role of family income and parental time investments for skill production (see Levy and Duncan 2000; Jenkins and Schluter 2002; Carneiro and Heckman 2002, 2003; Morris et al. 2005; Carneiro et al. 2010; Del Boca et al. 2014).

In contrast, there is no empirical evidence on complementarity in school inputs. Existing studies take account of self-productivity by allowing test scores in a stage to depend on test scores in the previous stage, but they do not usually allow for heterogeneity in the effect of school inputs across children with different levels of cognitive abilities (Hanushek 1997; Todd and Wolpin 2003; Rivkin et al. 2005; Jepsen and Rivkin 2009). There is descriptive evidence that has found higher returns to school inputs at higher ability levels using quantile regressions, hence allowing the effect of inputs to vary along the test score distribution (e.g. Eide and Showalter 1998; Rangvid 2007; Amini and Commander 2012). Figlio (1999) adopts a flexible education production model (relaxing the additivity and homotheticity assumptions) and shows that productivity of school inputs varies across different levels of student achievement as well as by level of other inputs. However, only Mueller (2013) uses an experiment, the Tennessee STAR experiment, which allows him to make causal inference as there is random variation in both school inputs (seniority of teachers and class size) and children's achievements. Mueller considers the effect of class size, teacher seniority and their interaction at different quantiles of student achievement. He finds that senior teachers generate class size effects which are most beneficial at the middle and higher end of the achievement distribution. None of these studies is however able to control for the endogeneity caused by unobserved non-cognitive or other unobserved child capabilities.

The remainder of this paper is organized as follows. Section 2 discusses estimation methods and the identification strategy used to produce our empirical evidence on self-productivity and complementarity. Section 3 describes the data, while Section 4 and 5 present the main results and robustness checks respectively. Finally, Section 6 concludes.

2 Methods

We focus on cognitive development from the end of primary school to the end of compulsory schooling, i.e. from about 11 to 16 years of age, and adopt the following education production model:

$$Y_{ih,16}^* = f(I_{ih}^F, I_{ih}^S, X_{ih}, Y_{ih,11}^*, \mu_{ih}), \quad (1)$$

where $Y_{ih,16}^*$ and $Y_{ih,11}^*$ are unobserved latent cognitive abilities of child i in family h at ages 16 and 11, I_{ih}^F is the family investment in child cognitive development between ages 11 and 16, I_{ih}^S is the corresponding school investment, X_{ih} is a row vector of other child, household and school characteristics, which are not direct investments in children’s cognitive skill but may affect it (e.g. gender, ethnicity, language spoken at home, free school meal eligibility, special educational needs, number of siblings, school characteristics and student composition), and μ_{ih} is the unobserved child time-invariant endowment which captures unobserved cognitive abilities as well as other unobserved child capabilities such as non-cognitive abilities and health.

Model (1) allows cognitive ability at age 16, $Y_{ih,16}^*$, to depend on cognitive ability at age 11, $Y_{ih,11}^*$, and we can test whether there is self-productivity of cognitive ability by testing whether $\frac{\partial f(\cdot)}{\partial Y_{ih,11}^*} > 0$. Furthermore, we consider two different specifications of model (1), one where the productivity of school inputs does not change across the level of lagged cognitive ability (homogenous effect) and another one where we allow for the interaction between school inputs and lagged abilities (heterogenous effect). We use the second specification to test the presence of complementarity, i.e. whether cognitive ability at age 11 makes school inputs more productive, $\frac{\partial^2 f(\cdot)}{\partial I_{ih}^S \partial Y_{ih,11}^*} > 0$.

We estimate model (1) by using the universe of students enrolled in state schools in England who took their age 16 school-leaving exams in the period 2007-2010. For this sample we are unable to observe family and school investments; but we can observe school expenditure per student, which we use as a measure of school inputs, and three measures of cognitive ability each at ages 11 and 16, which are test scores in Mathematics, English and Science obtained at the end of primary school and of compulsory schooling.

Our estimation strategy extends the two-step estimation used by Nicoletti and Rabe (2012) and Del Boca et al. (2012) to allow for a heterogenous effect of investments across

students with different levels of lagged cognitive abilities. The first step is a within-pupil between-subject estimation that allows us to control for the endogeneity of the lagged test caused by unobserved child specific endowments, while the second step is a sibling fixed effect estimation that allows us to control for the potential correlation between school expenditure and unobserved family inputs. In the following we first explain the method when the effect of school inputs is assumed to be homogenous and then extend the method to the case of a heterogenous effect.

2.1 Specification with homogenous effect of school investment

We assume that the relationship between each of the three test scores observed at age 11 and 16 and the unobserved latent cognitive skill at the corresponding age follows a classical measurement error model²

$$Y_{ihs,11} = Y_{ih,11}^* + e_{ihs,11} \text{ and } Y_{ihs,16} = Y_{ih,16}^* + e_{ihs,16}, \quad (2)$$

where the subscript s indicates the test subject and takes value 1 for Mathematics, 2 for English and 3 for Science, $e_{ihs,16}$ and $e_{ihs,11}$ are subject-specific random components identically and independently distributed across children, households and test subjects with mean zero and variance σ_e^2 , and are independent of the inputs in the production model and of the true latent skill at age 11 and 16, $Y_{ih,11}^*$ and $Y_{ih,16}^*$. The random components $e_{ihs,16}$ and $e_{ihs,11}$ measure the deviation of the subject specific skill from the general latent skill. We assume that there is no correlation between $e_{ihs,16}$ and $e_{ihs',11}$ if $s \neq s'$, but we allow for persistence in the subject-specific ability across age, i.e. $Cov(e_{ihs,16}, e_{ihs,11}) \neq 0$.³ Furthermore, we assume that the persistence in $Y_{ih,t}$ is identical to the persistence in $Y_{ih,t}^*$. More precisely we assume that the correlation between $Y_{ih,16}^*$ and $Y_{ih,11}^*$ net of the explanatory variables in the

²Imposing a classical measurement error model is equivalent to imposing a factor model with a single factor and equal factor loadings. The psychologist Spearman (1904) is the pioneer of factor analysis and was the first to apply it to capture a latent measure of skill which he called general intelligence or g-factor. But single factor models, to take account of measurement error in observed cognitive skill tests, have also been used more recently by economists (e.g. Cunha and Heckman 2008).

³Note that it is possible that there are spillover effects from one subject to the other that operate through mediating factors such as the level of confidence, attitudes toward school, etc. For example, high achievement in one subject may increase confidence or improve attitudes towards learning which may spill over to subsequent achievements in other subjects. We take account for these mediating factors by applying individual fixed effect estimation, therefore controlling for unobserved confidence levels, learning attitudes and other unobserved child-specific factors.

education production model is identical to the corresponding correlation between $Y_{ihs,16}$ and $Y_{ihs,11}$. This implies that the net correlation between $Y_{ih,16}^*$ and $Y_{ih,11}^*$ is also equal to the correlation between $e_{ihs,16}$ and $e_{ihs,11}$. In our robustness checks we provide evidence that this assumption holds empirically.

Under the assumptions defined above and imposing that the production function (1) is additive, separable and linear in its arguments, we can rewrite it as

$$Y_{ihs,16} = \alpha + I_{ih}^F \beta_F + I_{ih}^S \beta_S + X_{ih} \gamma + Y_{ih,11}^* \rho + \mu_{ih} + e_{ihs,16}, \quad (3)$$

where we replaced the unobserved latent cognitive skill at age 16 with the observed test score in subject s and $s = 1, 2, 3$ and where α is the intercept, β_F and β_S are scalar parameters capturing the effects of family and school disinvestments, γ is a column vector with elements given by the effects of the explanatory variables X_{ih} , and ρ is a scalar parameter measuring the persistence in cognitive ability. Model (3) is usually known as the value added model (see Todd and Wolpin 2003) and it has been extensively used in previous empirical papers to evaluate the contributions of school inputs in a specific stage of the child's school life by controlling for the child's cognitive skill at the beginning of the stage (see Hanushek 1997; Meghir and Rivkin 2011; Holmlund et al. 2010).

Since the lagged cognitive skill $Y_{ih,11}^*$ is unobserved, we replace it with the lagged test in subject s and rewrite equation (3) as

$$Y_{ihs,16} = \alpha + I_{ih}^F \beta_F + I_{ih}^S \beta_S + X_{ih} \gamma + Y_{ihs,11} \rho + \mu_{ih} + u_{ihs,16}, \quad (4)$$

where $u_{ihs,16} = e_{ihs,16} - e_{ihs,11} \rho$. We provide details of the asymptotic bias of the ordinary least squares estimator in Appendix A.

To take account of the endogeneity of the lagged test score caused by the unobserved child-specific endowment, μ_{ih} , we adopt a *two-step estimation*. In the first step we use test scores in three subjects and three corresponding lagged test scores for each child to estimate a *child fixed effect model*. This allows us to control for the unobserved child specific endowment that is invariant across subjects and to consistently estimate ρ in the value added model (4). However, this estimation is unable to identify the return to school investment because this does not vary across the three tests.

Therefore we implement a second step where we use the estimated coefficient ρ to compute a new dependent variable ($Y_{ihs,16} - Y_{ihs,11}\rho$) which we regress on the remaining variables,

$$Y_{ihs,16} - Y_{ihs,11}\rho = \alpha + I_{ih}^F\beta_F + I_{ih}^S\beta_S + X_{ih}\gamma + \mu_{ih} + u_{ihs,16}. \quad (5)$$

We estimate this regression using *sibling fixed effect estimation* to control for potential unobserved variables that do not vary between siblings and in particular to control for the family investment I_{ih}^F which we are unable to observe in our sample (see Rosenzweig and Wolpin 1994; Altonji and Dunn 1996; Behrman et al. 1996; Todd and Wolpin 2007). We discuss the credibility of the assumptions imposed in our model in more detail in Section 2.3.

2.2 Specification with heterogenous effect of school investment

In model (3) the return to school investment does not vary across different levels of past achievement, i.e.

$$\frac{\partial Y_{ih,16}^*}{\partial I_{ih}^S} = \beta_S. \quad (6)$$

We are interested in testing whether the return to school investments between age 11 and 16 depends on the level of skills observed at a previous stage (age 11). Therefore we add to model (3) an interaction term between school investment and lagged cognitive ability, ($I_{ih}^S Y_{ih,11}^*$),

$$Y_{ihs,16} = \alpha + I_{ih}^F\beta_F + I_{ih}^S\beta_S + (I_{ih}^S Y_{ih,11}^*)\beta_{SY} + X_{ih}\gamma + Y_{ih,11}^*\rho + \mu_{ih} + e_{ihs,16}, \quad (7)$$

i.e. by assuming that the marginal effect of school investment is

$$\frac{\partial Y_{ih,16}^*}{\partial I_{ih}^S} = \beta_S + Y_{hs,11}^*\beta_{SY}. \quad (8)$$

Notice however that model (7) imposes the marginal effect of school investment to be linear in the lagged test. To relax this linearity assumption in our empirical application we allow the marginal effect of school investment to vary across deciles of the lagged test score, i.e. we assume that

$$\frac{\partial Y_{ih,16}^*}{\partial I_{ih}^S} = \beta_{S,10} + \mathbf{I}(d_{j-1}^* < Y_{ih,11}^* \leq d_j^*)\beta_{S,j}, \quad (9)$$

where $\mathbf{I}(d_{j-1}^* < Y_{ih,11}^* \leq d_j^*)$ is an indicator function, d_j^* is the j -th decile of $Y_{ih,11}^*$ ($j=1,\dots,9$), d_0^* is the minimum value taken by $Y_{ih,11}^*$, $\beta_{S,10}$ is the effect of school inputs for children whose

lagged test is in the top decile, and $\beta_{S,j}$ is the differential effect of the school inputs for children whose lagged test is in between the $(j - 1)$ -th and the j -th decile. This implies adopting the following model,

$$Y_{ihs,16} = \alpha + I_{ih}^F \beta_F + I_{ih}^S \beta_{S,10} + \sum_{j=1}^9 I_{ih}^S \mathbf{I}(d_{j-1}^* < Y_{ih,11}^* \leq d_j^*) \beta_{S,j} + X_{ih} \gamma + Y_{ih,11}^* \rho + \mu_{ih} + e_{ihs,16}. \quad (10)$$

To estimate model (10) we proceed again with a two-step estimation. In the first step we use child fixed effect estimation to estimate the persistence, ρ , and the differential effect of the school investment at the first 9 deciles of the lagged test distribution, $\beta_{S,j}$ for $j = 1, \dots, 9$. In the second step we use sibling fixed effect estimation to estimate the baseline effect, i.e. the effect of school inputs at the 10th decile of the lagged test distribution. We provide more details on the estimation of model (10) and on the assumptions imposed for its consistency in Appendix B.

2.3 Identification of complementarity

As stressed in the Introduction, inference on complementarity requires exogenous variation in both investments and lagged abilities. We provide details on how we use exogenous variation in both lagged cognitive abilities and school investments in the next two sub-sections.

2.3.1 Exogenous variation in lagged cognitive ability

Past cognitive ability can depend on unobserved dimensions of child ability, such as non-cognitive ability and health, and on unobserved family inputs. This implies that part of the variation in lagged cognitive ability is endogenous (see Todd and Wolpin 2003 and 2007; Andrabi et al. 2011). To control for the potential endogeneity issue, we use child fixed effects methods in the first step and sibling fixed effect methods in the second step of our estimation procedure. The first step provides consistent estimation of the persistence ρ and the interaction coefficients $\beta_{S,j}$ ($j = 1, \dots, 9$) in model (10) even in the presence of unobserved student specific capabilities, as long as they have the same effect on the return to school inputs across subjects. Our estimates are also consistent if parental investments affect the return to school inputs, as long as these investments do not differ or have a differential effect across

subjects. By using child fixed effects estimation we are purging our estimates of subject-invariant endogeneity, but it may be that there is remaining subject-specific endogeneity affecting our results. For example, a student may have a special talent in Maths that allows her to improve rapidly in Maths over time, and this improvement could be wrongly attributed to dynamic complementarity. We test this possibility in various ways, for example by investigating whether groups of student specialise in particular subjects and by excluding from our estimates groups that we suspect could be affected by “learning anomalies” in particular subjects. These tests indicate that dynamic complementarity holds regardless of any subject-specific endogeneity. In summary, the first-step estimation allows us to make consistent inference on the presence of self-productivity and complementarity, i.e. on whether $\rho > 0$ and whether $\beta_{S,j}$ is increasing in j in model (10).

While our inference on the presence of self-productivity and complementarity is consistent in the presence of subject-invariant unobserved child capabilities or parental investments, the estimation of the baseline effect of school inputs is not necessarily consistent. By baseline effect of school inputs we mean the effect of school inputs at the 10th decile of lagged cognitive ability, i.e. $\beta_{S,10}$ in model (10). The baseline effect is not a main parameter of interest in this paper, as we can investigate whether or not the return to school inputs is increasing in cognitive ability without a consistent estimate of the baseline effect. We discuss consistency because the baseline effect may be of substantive interest in other applications.

The baseline effect is estimated in the second step of our procedure using sibling fixed effects estimation and it is consistent only if either siblings share the same family characteristics, or differences between siblings in family characteristics, such as potential differences in family investment between siblings, are uncorrelated with differences in school inputs (after controlling for differences in the other explanatory variables in our production model).

A possible situation in which there is an association between sibling differences in cognitive ability and in school inputs is the case where parents decide to compensate or reinforce these differences, for example by enrolling their children in different schools or through differential parental investments (Behrman et al. 1982). By including in our estimates only siblings going to the same school we avoid the bias resulting from differential investments that operate through school choice. We take into account that parental investments might compensate or reinforce for the sibling difference in cognitive ability by controlling for test

scores observed at the end of primary school (lagged test scores), i.e. by allowing the sibling difference in parental investments to depend on the observed difference in their lagged test. Arguably lagged test scores are the main measures of child cognitive abilities that parents can observe to decide how to invest in their children. Note that we also control for various other child characteristics that may induce parents to invest differentially in siblings, including birth order, birth month, having special educational needs (learning difficulties that are related to behavioral or health issues), and being identified by the school as gifted or talented.⁴ Parents may also vary their inputs by school expenditure. However, the variation in school expenditure we exploit is based on idiosyncratic changes within one school between cohorts, and we suspect that these changes are difficult to be perceived by parents.

2.3.2 Exogenous variation in school investment

Differences in school investments across state schools are generally not random because the allocation of public funding tends to favor schools with a higher fraction of students from disadvantaged backgrounds. We take account of the redistributive nature of school inputs by controlling for the school characteristics that are used to allocate resources (for example the proportion of students on free school meals). After controlling for these characteristics, we have exogenous variation in expenditure per student across schools. This is because in the time-period considered in our empirical analysis (2005-2010) there was a substantial real increase in funding per student from an average of 4,690 pounds in 2005 to 5,750 pounds in 2010 (23% increase in 2010 prices).⁵ Moreover, the rules used to allocate funding across schools vary regionally and have built-in delays in adapting to changes in school characteristics so that similar schools can receive very different funding levels.

Most funding for state schools in England comes from central government which hands money to local education authorities, of which there are 154. The central government grant

⁴Another situation that can lead to a biased estimator of the baseline effect is the case when one of the siblings has serious learning difficulties, physical disabilities or significant behavioral problems. These children are usually enrolled in so-called “special schools” which provide specific support for children with more extreme needs and which typically have higher resources. To avoid the resulting bias, we exclude these types of schools from our sample.

⁵In our empirical analysis we consider test scores of four cohorts of students, taking exams in 2007, 2008, 2009 and 2010. School inputs are three-year averages of expenditure per student, so that for a student taking exams in 2007, inputs will be from the period 2005-2007.

is calculated using a funding formula mostly based on student numbers, educational disadvantage and area costs. However, a so-called spend-plus methodology is applied whereby local authority grants are determined as flat-rate increases on the grant received the previous year - with a historical starting point in 2005-06 - plus an extra increase based on the formula. “So, current levels of school funding are based on an assessment of needs which is out of date, and on historic decisions about levels of funding which may or may not reflect precisely what schools needed then” (Department for Education 2011, p. 3).

Local authorities then use their own funding formulas to hand out the money received from central government to schools. However, a major constraint that local authorities face is the Minimum Funding Guarantee, introduced in 2004-05, which guarantees each school a minimum increase per student per year. In 2010-11 half of the annual increase in funding was used to meet the Minimum Funding Guarantee (Chowdry and Sibieta 2011), so it largely limits the freedom with which local authorities can choose their funding rules (Levačič 2008). Apart from student numbers, many local authorities allow more funding for students from deprived backgrounds (eligible for free school meals), with special educational needs and with English as an additional language (Chowdry and Sibieta 2011). There is considerable variation between local authorities in the formula used (West 2009).

The combination of spend-plus methodology and Minimum Funding Guarantee has weakened the relationship between school funding levels and educational need. In 2010-11 7% of secondary schools had a level of funding at least 10% lower than predicted using observable characteristics, and 6% had funding at least 10% higher (Chowdry and Sibieta 2011, p. 12).

To illustrate the variation in expenditure we exploit in this paper, Table 1 gives a preview of the between-sibling variation in expenditure per student in our sample (see the next section for details on the data). Table 1 shows that the majority of siblings in our data (85%) go to the same school. Among siblings going to the same school, on average the younger sibling in a family will have received an annual input which was 349 pounds higher than that received by the older sibling in the same school, with a standard deviation of 283 pounds. We know that the expenditure per student has increased over our sample years, so we want to make sure the between-sibling variation is not driven only by a time trend. Therefore we also present the mean and standard deviation of the expenditure differential between the younger and older sibling a) net of a time trend (we control for academic year in our models), and b)

net of a school-specific time trend to check for local trends. The Table shows that the mean differences in expenditure per student net of the time trend and the school-specific time trend is indeed lower, but reassuringly the standard deviation is still substantial. We can also see that a minority of siblings attend different schools. The mean between-sibling differences are smaller for this group, but the standard deviation is larger - both for the overall expenditure and for expenditure per student net of the trends. As we cannot assume for these siblings that differences in expenditure are exogenous, as they could reflect parental school choice associated with differential sibling ability, we restrict our empirical analysis to siblings going to the same school.

3 Data

We use the National Pupil Database (NPD), which is available from the English Department for Education and has been widely used for education research. The NPD is a longitudinal register dataset for all children in state schools in England, covering roughly 93% of students. It combines student level attainment data throughout primary and secondary school with student characteristics.⁶

In England full-time education is compulsory for all children aged between 5 and 16. The education during these years is divided into four Key Stages and at the end of each stage (at ages 7, 11, 14 and 16) there is an assessment of the student's educational achievements. Assessments at the end of Key Stages 2 and 4 are externally marked and therefore comparable across different schools. Key Stage 2 tests are taken at the end of primary school, usually at age 11, in the core subjects of English, Mathematics and Science. Key Stage 4 tests are taken at age 16 at the end of compulsory schooling, and are either the General Certificate of Secondary Education (GCSE) exams or equivalent vocational or occupational exams. Students decide which GCSE courses to take, and because English, Mathematics and Science are compulsory study subjects, virtually all students take GCSE examinations in these topics, plus others of their choice, with a total of ten different subjects normally taken. In addition to GCSE examinations, a student's final grade may also incorporate coursework elements.

⁶The NPD also holds attainment data for pupils in non-maintained and independent schools, but we do not consider them in our sample because we cannot observe their background characteristics.

Student-level variables

Our outcomes of interest are GCSE or equivalent vocational test results at age 16 in English, Mathematics and Science. Students receive a grade for each GCSE course, where pass grades include A*, A, B, C, D, E, F, G. We transform these grades into a continuous point score which we refer to as the Key Stage 4 score.⁷ We control for lagged cognitive skills using Key Stage 2 continuous test scores in English, Mathematics and Science. All test scores are standardized to have a mean of zero and a standard deviation of one.

Individual and family background variables available in the NPD include the gender of the student, ethnicity (white British, black, mixed, Indian, Pakistani and Bangladeshi, Chinese), whether or not the first language spoken at home is English, whether special educational needs have been identified for the child and whether the student has been classified as gifted and/or talented. Moreover, we can identify whether or not a student is eligible for free school meals (FSM). FSM eligibility is linked to parents' receipt of means-tested benefits such as income support and income-based job seeker's allowance. We also include the number of months a student is older than an August-born (the youngest in a school cohort), to control for relative age at test within the cohort, and for the Income Deprivation Affecting Children Index, which is a measure of deprivation in the children's residential neighborhood available in the NPD.

Finally we consider the number of siblings to control for the size of the household and an indicator variable for being the oldest pupil in a household or a singleton to control for possible birth-order effects. Siblings are defined as co-residential students, and we identify them by matching address data released under special conditions. The first year that full address details were collected in the NPD across all student cohorts was 2007. Siblings are therefore defined as students in state schools aged 4-16 and living together at the same address in January 2007.⁸ Siblings who are not school-age, those in independent schools and those living at different addresses in January 2007 are excluded from our sibling definition. Step and half siblings are included if they live at the same address, and we are not able to distinguish them from biological siblings.

⁷We use a scoring system developed by the Qualifications and Curriculum Authority which assigns 16 points to pass grade G, and 6 points are added for each unit improvement from grade G.

⁸We have address data for 2007 only so cannot update the sibling status across years.

School-level variables

We merge school-level expenditure information from Consistent Financial Reporting data sets for 2004-2010 to the NPD. These data sets contain details on different types of income and expenditure for each school, separately for each academic year. The data allow us to derive the expenditure per student which excludes capital expenditure such as new construction, but includes expenditure items such as learning resources which may benefit students for several years. For each student taking his/her key stage 4 exams in the academic year t , the school expenditure per student is computed as the expenditure per student in his/her school expressed in 2010 prices using the DGP deflator and averaged over the years $t - 2$, $t - 1$ and t .⁹

In addition we add school-level characteristics to the NPD using Schools, Pupils and their Characteristics tables published by the Department for Education. These tables are derived from annual school censuses. School-level characteristics include the number of students in the school and variables describing the student composition, in particular the proportion of students that receive free school meals, whose first language is English, that are of white British, black, mixed, Indian, Pakistani/Bangladeshi and Chinese ethnicity and that have special educational needs (with and without statements). Again we average these variables describing the student composition over three years. We also add cohort mean Key Stage 2 test scores in English, Science and Maths as school-level controls for prior attainment within the school. Academic year dummies are included to control for trends in test results (“test inflation”) and for trends in expenditure.

Estimation sample

For our analysis we select a sample of the oldest two siblings going to the same school from each household. Focusing on two siblings from each family avoids having to expand the dataset to include all sibling pair combinations within each household with the risk of over-representing households with a large number of children. In the vast majority (95.5%) of families there are only two siblings observed taking Key Stage 4 exams in the observation period. We remove twins from the data set as well as siblings who begin school in the same year as expenditure will not vary between them. The sample includes all students that took Key Stage 4 exams in 2007 or in one of the three following years (2008, 2009, 2010).

⁹We consider alternative ways of measuring school expenditure in the Robustness section.

We remove students with duplicate data entries or with missing data on any of the background or school-level variables from the dataset. Moreover, we retain only students for whom we have non-missing test scores for all outcomes at both Key Stages 2 and 4 which leads to a reduction in sample size of 13%.¹⁰ We also exclude “special schools” that exclusively cater for children with specific needs, for example because of physical disabilities or learning difficulties, as well as schools specifically for children with emotional and/or behavioral difficulties. Academy schools which have been introduced in 2000 and allow schools more autonomy and flexible governance are also excluded. In 2007 about 1.5% of schools had academy status, rising to about 6.5% in 2010. Finally, we exclude the top 1% of the expenditure per student distribution to remove extreme outliers from the data set. The remaining sample contains 359,470 students (179,735 sibling pairs).

Table 2 describes our sample. It displays individual characteristics in the top panel and school characteristics in the bottom panel. 92% of students have English as their first language, 18% are classified as gifted and talented and 16% have special educational needs. The proportion of White British students is 87% and 10% of students in our sample are eligible to receive free school meals. The bottom panel of Table 2 shows that secondary schools are quite large with more than 1,000 students in a school on average. The school-level proportions of students with free school meal eligibility, ethnicity and English as their first language are comparable to the individual level means.

4 Empirical results

In Table 3 we report the main estimation results of the education production model. All results refer to value added models estimated using the sample of siblings going to the same school and pooling together observations on the tests at age 16 in Mathematics, English and Science, i.e. imposing the same model coefficients across subjects. Note that this pooled sample has three times the number of observations reported in the descriptive statistics, Table 2. The explanatory variables include the lagged test score at age 11, school expenditure per student, the set of child and school characteristics described in Table 2 and dummies for

¹⁰Missing cases are concentrated among low attaining students that are more likely to be absent at exams or, at Key Stage 4, choose not to take exams in one or more of the core subjects. Comparing the original with the retained sample the average test score is reduced by about 1%

the academic year to control for grade inflation, trends in funding levels and other potential changes over time.

In columns (1) to (3) of Table 3 we compare results based on three different estimations of the value added model: (i) ordinary least squares estimation (OLS), which neglects unobserved family and school characteristics and the endogeneity of lagged cognitive ability, (ii) sibling fixed effect estimation which controls for unobserved family and school characteristics and inputs that are invariant across siblings, but does not control for the endogeneity of lagged test scores, (iii) two-step estimation (two-step sibling fixed effect) which controls for both unobserved family and school heterogeneity and endogeneity of the lagged test. Note that unobserved school characteristics that are persistent over time are controlled for in the sibling fixed effect and two-step estimation because our sample considers only siblings going to the same school. We report the results of these three estimation methods for the model that allows the effect of expenditure per student to vary across different deciles of the lagged cognitive ability (see top panel of Table 3) and for the model that imposes equal return to the expenditure per student across the distribution of the lagged cognitive ability (see bottom panel of Table 3).

Our preferred estimation is the two-step estimation of the value added model with heterogeneous effects of the expenditure per student (see column 3 in the top panel of Table 3), which we use as benchmark for comparison.¹¹ These benchmark estimates provide evidence of the presence of strong complementarity and self-productivity of cognitive ability. A £1,000 increase in the expenditure per student has a statistically significant effect of 3% of a standard deviation for children at the bottom of the lagged test score distribution, while it leads to an increase of 9% of a standard deviation in test scores for children at the top of the lagged test score distribution.

Although the overall effects of expenditure per student are modest, the differences by lagged ability are sizeable: productivity of expenditure is three times higher for students at the top of the lagged attainment distribution than for those at the bottom. The returns to school expenditure are strictly monotonically increasing in the lagged test score and the differences between the students in two adjacent deciles are always statistically significantly

¹¹We do not bootstrap the standard errors to take account of the fact that we replace ρ and $\beta_{S,j}$ for $j = 1, \dots, 9$ with their estimates because our first step has very low standard errors and makes use of the universe of pupils, so we do not expect our standard errors to change much.

different from zero at the 5% level. We attribute the increasing effect of school expenditure by level of lagged test score to complementarity between school inputs and past ability. One situation we can think of that would lead to such results in absence of complementarity is a situation where within a school more resources are invested into children with higher lagged ability (e.g. teachers give more attention to brighter students). But this seems to contradict existing school policies, which are targeted to provide more school resources and support to children who are low achievers. We discuss further possibilities in the robustness section.

The persistence ρ is the coefficient that captures the intensity of self-productivity. Table 3 shows that an increase of one standard deviation in the test score at the end of primary school leads to an increase of about 22% of a standard deviation in the test score at the end of compulsory schooling (see parameter ρ in the top panel). This suggests that early school interventions can be more effective than later ones because they not only directly increase cognitive ability at the time but also future cognitive ability. This is because higher cognitive ability attained at one stage both persists into the next stage and increases the return to school inputs in later stages through a multiplier process.

We can see that the models based on OLS and sibling fixed effects methods (columns 1 and 2) underestimate the effect of expenditure per student and overestimate the persistence in cognitive ability. This is true for both the models with homogenous and heterogenous effect of school expenditure (see bottom and top panels of Table 3). These models do not control for the endogeneity of the lagged test nor, in the case of the OLS model, for unobserved family and school characteristics. Since these omitted variables are likely to be positively correlated with the lagged test score net of the remaining control variables in the production model, the overestimation of the persistence is in line with the asymptotic bias we expect (see Appendix A). We also expect that an overestimation of the persistence may lead to a underestimation of the input effect in the production function and indeed this is what we find empirically.

Imposing the same return to expenditure per student across the lagged cognitive ability distribution causes additional overestimation bias of the persistence (see ρ coefficients in the bottom panel), while the estimated homogenous effect of expenditure per student seems close to the corresponding estimated effect at the 5th decile of the lagged test distribution. This overestimation of the persistence is in line with the expected asymptotic bias caused by

the omission of the interaction terms between school investment and dummy variables for the different deciles of the lagged test.¹²

To check for potential differences by gender in the learning process, we also estimate the two-step education production model separately by gender. For these estimations we select, respectively, all sister and brother pairs in the sample and implement the two-step estimation with same-sex sibling fixed effects in the second step. Overall, the separate results for boys and girls show that the effect of school expenditure increases monotonically with lagged test scores and that dynamic complementarity holds for both boys and girls.

The gender comparison is quite interesting; while the persistence in the cognitive ability is almost identical for boys and girls at about 0.22, the productivity of the expenditure per student increases more steeply with the level of lagged cognitive ability for boys than for girls. For example, for boys in the bottom decile of lagged cognitive ability an increase of £1,000 in the expenditure per student leads to an increase of 4.2% of a standard deviation in cognitive ability, whereas this increase rises to 11.6% for boys at the top decile. The corresponding percentages for girls are 3.4% and 7.8%, and the differences are statistically significant.

While there is a literature on gender gaps in productivity, there are no papers that look at explaining gender differences in productivity along the distribution of cognitive ability. Similarly to previous empirical papers on gender gaps in test scores, we find that boys throughout the distribution perform worse than girls both at the end of primary and at the end of compulsory schooling. Our results also suggest that there is a partial catch up of boys with girls, but only for boys who are at the higher end of the test score distribution.

¹²More specifically we expect to have overestimation of the persistence if the return to school inputs increases when moving from the bottom to the the top decile of the lagged test distribution and if the omitted interaction terms are negatively correlated with the lagged test after controlling for the remaining variables in the production model.

5 Robustness checks

5.1 Threats to identification of dynamic complementarity

Our preferred estimation indicates that the effect of expenditure per student is increasing across the distribution of lagged cognitive ability, but to confidently attribute this to complementarity we need to check that the higher return to school expenditure for children with higher lagged cognitive ability is not caused by subject-specific endogeneity.

Our estimates are purged of a child fixed effect, but there could be parental inputs, study efforts and child traits that are subject specific, therefore affecting test score gains in one subject but not the other. We discuss here the possibility of (1) subject specialisation (i.e. study effort that differs across subjects), (2) subject specific parental investments, (3) the existence of multiple dimensions of skills that differentially affect tests in different subjects.

Specialization in a subject may bias our results because students who have preferences for a specific subject may be more likely to gradually increase their effort in this subject with respect to others. This increasing effort may lead to a high test score in the preferred subject at the end of primary school and to an even higher test score at the end of compulsory schooling, so that the strength of complementarity could be overestimated. This could be the case for boys, for example, who are more likely to prefer Mathematics and to obtain higher scores in Mathematics than in English and Science and for whom we observe a higher level of complementarity than for girls (see Table 4). However, we find that the average difference between a student's highest test score and the remaining two test scores does not increase between age 11 and 16 for boys or for girls, indicating that there is not a strong subject specialization. Furthermore, the increase in the average standardized test score in Mathematics between age 11 and 16 is smaller for boys than for girls. Similarly we also check whether specialization can be found for other groups, including by ethnicity, free school meal status and special educational needs. The average difference between the best and the other two subject decreases over time for all groups, giving no evidence of specialization.

It is possible that parental investments reinforce differences in skills between subjects. Suppose for example that parents are better able to help in one subject than another, or they have a preference for a particular subject and therefore invest in it. This would again

lead to specialization, which could bias our results on dynamic complementarity. But, as stated above, the existence of such specialization is not confirmed empirically for any group of students. On the other hand, parental investments could be allocated so as to compensate for low skills in a specific subject. In this case parental investments would attenuate the degree of complementarity we estimate. Compensating parental investments therefore imply that, if anything, we underestimate dynamic complementarity.

We are also concerned about multiple dimensions of skills that may differentially affect tests in different subjects, causing subject-specific endogeneity. Examples of such skills are non-cognitive skills and health. To give an example of a health condition, a student diagnosed with dyslexia will likely be affected in her learning in English but not Maths, whereas a student suffering from dyscalculia (difficulty in learning and comprehending arithmetics) would by expectation be affected in their Maths but not English scores. These subject specific skills would not be eliminated in our estimation and bias our estimate of dynamic complementarity.

We are of course unable to observe all the specific conditions and abilities that may differently affect students' test scores across subjects. Our robustness check consists in generously excluding from our sample all students (40.4% in total) that may *conceivably* be affected by subject-specific endogeneity and to establish that dynamic complementarity holds for the remaining sample. Specifically, we can identify in our data three groups of students among which we would expect subject specific endogeneity to be concentrated. These are children with (i) special educational needs, (ii) gifts or talents, and (iii) behavioural issues.¹³ The special educational needs category signposts children with learning difficulties that could be subject specific (e.g. dyslexia and dyscalculia). Children that have been identified by the school as being gifted or talented by definition have a special ability in one or more subjects. Finally, we do not know whether behavioral issues affect learning differentially across subjects, and we want to allow for that possibility. In column (2) of Table 5 we report estimates for the sub-sample of students that have never been flagged up by the school

¹³16.5% of students in our sample are classified as having special educational needs, and 18.5% of the sample have been identified by the school as being gifted or talented. Finally, we can observe whether students have been excluded by the school for fixed amounts of time (from a session up to several days) because of bad behavior. This could be because of disruptive behavior in class or violence, for example. 11.7% of our sample of students gets excluded at least once during their final year at school or the year before, and we take this as a marker of behavioral issues.

as having issues, good or bad, and we find that there is still strong evidence for dynamic complementarity, although the difference in return to school expenditure between the top and bottom decile of the lagged test distribution is attenuated.

There is another way we can assess the potential differential effect across subjects of unobserved inputs such as parental investments, child effort or unobserved skills. This is a robustness check where we estimate our model based on Maths and Science tests only and compare results to the model estimated across the three subjects. Here we assume that test scores in Maths and Science both measure analytic abilities that are quite similar, whereas the English test score measures a different type of ability. If Maths and Science have a similar underlying ability we expect unobserved child characteristics to affect Maths and Science scores in a similar way so that the issue of subject specific endogeneity should cancel out. Table 5, column (3) shows the results of this exercise. The baseline effect of expenditure per student is very similar to that estimated on the sample including test scores in all three subjects, and the degree of complementarity is also very similar. In other words, even if we base our inference on test scores on subjects for which we expect a similar effect of parental investments, child effort and child unobserved abilities, our inference on dynamic complementarity remains valid.

5.2 Measurement error in school test scores

Measurement error in test scores can bias the estimation of skill production models (see Cunha and Heckman 2008; Cunha et al. 2010; Cunha 2011) and the extent of such measurement error and resulting bias depend on the specific test score used. Cunha et al. (2010) find that 95% of the variance in the the Motor and Social Developments Score at year of birth is explained by measurement error rather than true signal; in contrast, only 4% of the variance in the PIAT Reading Recognition Scores at ages 5-6 is due to measurement error.¹⁴

Our estimation procedure explicitly allows for measurement error in test scores and it is consistent under the assumptions of the classical measurement error model (2) and of equal persistence in the three test scores (or equivalently of equal persistence in the true latent ability and in the subject specific test score). To test the assumptions imposed by model

¹⁴See Table 2A in Cunha et al (2010) for percentage of the variance due to measurement error for other tests.

(2), we consider a factor analysis of the test scores in Mathematics, English and Science at age 16 using the NPD data and check whether the hypothesis of a single factor model with equal factor loadings is valid. We find that the first factor explains more than 80% of the total variance, the loadings of this factor for the three test scores are higher than 0.9, and the percentage of the variance of the test score unexplained by the first common factor (i.e. explained by error rather than true signal) is 14.3%, 12.7% and 18.5% for Maths, English and Science respectively. These results suggest that (i) the variance of the error term with respect to the variance of the latent cognitive ability is reassuringly small; (ii) the assumption imposed by the measurement error model (2) of a single factor model with identical factor loadings across subjects is credible.

Furthermore, the correlations between test scores in Maths, Science and English at age 11 and at age 16, which are 0.767, 0.674 and 0.706 respectively, confirm empirically that the assumption of equal persistence in the three test scores and in the true latent cognitive ability is satisfied.¹⁵

5.3 Relating test scores and school spending to real-life outcomes

In this section we address the issue that (i) our measures of achievement (test scores) have no natural metric so that results are potentially sensitive to arbitrary scaling of test scores, and (ii) our measure of school investments (spending) has no direct link to actual educational inputs. We do this by replacing each measure with variables that capture real-life outcomes. Our aim is to make sure that dynamic complementarity and self-productivity hold for models estimated using these alternative, real-life, measures.

First we address the issue that test scores are arbitrarily scaled and any monotonic transformation of the test score is also a valid scale, potentially leading to problems in value added models in particular because these compare (arbitrarily scaled) test scores over time (see Bond and Lang 2013). One solution to this problem suggested in the literature is to re-scale the test score by anchoring it to a real-life outcome (see Cunha and Heckman 2008; Cunha et al. 2010). In our data we are able to observe whether students stay on in school

¹⁵More details on the factor analysis and on the correlations between tests scores are reported in Nicoletti and Rabe (2012).

after completing compulsory education at age 16. Post-compulsory schooling prepares students for their A-levels which are required for University entry, and participation is therefore an important outcome. We rescale our test score by considering the probability to stay in schooling post 16. Specifically, we use non-parametric estimation to predict the probability to stay in education associated with each level of the test score by subject and Key Stage and we re-scale each of the possible values of a test score in terms of probability percentage points.

In column (2) of Table 6 we report the estimation results when using the rescaled (past and current) test scores.¹⁶ As we can see, there is a statistically significant increase in the return to expenditure from the 1st to the 6th decile, while the pattern of returns to expenditure from the 7th to the 10th decile is more erratic. This is likely the case because our outcome is binary, and most students at the higher end of the attainment distribution participate in post-compulsory schooling (64% of the students in our sample do). We therefore conclude that dynamic complementarity holds also when relating test-scores to a real-life outcome, but the outcome we can observe is too blunt to capture this along the whole attainment distribution. The results suggests that scaling of test scores does not reverse results in our application, and we use test scores for our benchmark results.

Our second issue is that we measure school investments using overall expenditure per pupil, but it is not immediately clear how spending relates to actual educational inputs. To give some sense of the effect of actual school inputs we use data on expenditure categories to divide the total expenditure into spending on teachers and on everything else (other) and include these two variables in our model. Column (3) in Table 6 shows the results. We find statistically and substantially strong differences in the return to expenditure on teachers across deciles of the lagged test, whereas there is no statistically significant difference in the return to other expenditure across deciles. These results suggest that there is strong complementarity between past abilities and school inputs that directly affect the student learning process, but not between past abilities and school inputs that are auxiliary to the learning process (such as learning resources, like books and computers).

¹⁶At the top decile a £1,000 increase in school expenditure implies an increase of 1.9 percentage points in the probability of staying in education at 16 - a substantially small effect.

Note, however, that school-level decisions on how to use extra funds are potentially endogenous here. The problem is alleviated to an extent by using sibling fixed effects estimation on siblings going to the same school in the second step of our estimation procedure which amounts to using a school fixed effect. However, in the first step there may be endogeneity of how schools allocate funds to different purposes, and we therefore prefer to measure inputs more generically using total expenditure per pupil.

Regarding measurement of school investments we present one final robustness check. For our benchmark estimates we use average per student expenditure across the three last school years as measure of school inputs. Using such an average implicitly assumes that the effect of expenditure does not diminish the larger the distance from the exam. It could also be argued that the correct measure of expenditure should include all the years between ages 11 and 16. For the purposes of this paper we are mainly concerned to check that our inference on dynamic complementarity holds for alternative ways of measuring expenditure. Because year-on-year expenditure is highly correlated, we can not jointly identify yearly expenditure effects on test scores. Instead we estimate separate models for each of the yearly expenditures, and we display the results for the expenditure in the last school year in column (4) of Table 6. We find that the expenditure effect reduces when using yearly expenditure rather than the 3-year average expenditure, and this can likely be caused by measurement error inherent in short term expenditure. Importantly for us, the results on self-productivity and dynamic complementarity remain qualitatively the same for each yearly expenditure measure (only last displayed).

5.4 Heterogeneity along other dimensions

Our empirical model allows for heterogeneity of the effect of school inputs by lagged achievement to allow us to test for dynamic complementarity. It is conceivable that heterogeneity could arise along other dimensions, and neglecting to allow for this could lead to model mis-specification. In particular, self-productivity and dynamic complementarity could be different by subject, for high and low spending schools, and when letting the persistence in test scores to be non linear. To see whether relaxing these restrictions affects our inference on dynamic complementarity, we allow both the returns to school expenditure and the

persistence to vary across subjects, level of school expenditure and decile of the lagged test score.

In columns (1)-(3) of Table 7 we report the estimation results when allowing for heterogeneity across subjects which we estimate using subject interactions. There are statistically significant differences in the ρ coefficients and in the returns to expenditure at different deciles of the lagged test across subjects, but these differences are not large and can be caused by slightly larger measurement error in test scores in Science compared to English and in English compared to Mathematics. The measurement error seems to lead to an attenuation bias for the persistence and to an overestimation of the return to expenditure per pupil at the top decile for Science and English. However, dynamic complementarity is still strong and the differences in the return to school expenditure between the top and the other deciles have a very similar size and pattern across subjects.

In column (4) and (5) of Table 7 we report the results when allowing for heterogeneity in the parameters between schools with a level of expenditure per student below and above the median. There are no substantial differences in the persistence and returns to school expenditure by level of expenditure and we cannot reject the equality of the coefficients between the two levels of expenditure at standard levels of significance.

Another assumption imposed by our benchmark model is the linearity in the relationship between current and lagged test score. Although both tests are measured in terms of standard deviations, the assumption of linearity may be too strong. For this reason we extend our benchmark model to allow the ρ coefficient to vary across deciles of the lagged test (not displayed). We find no clear pattern of the ρ coefficient across deciles and even if the differences across deciles are statistically significant, they are substantially very small.¹⁷ Furthermore, the evidence on dynamic complementarity remains similar to our benchmark results.

In conclusion, our benchmark results on dynamic complementarity are robust to different model specifications and largely unaffected by the assumptions imposed by our preferred model.

¹⁷Results are available from the authors upon request.

6 Conclusions

In this paper, we use register data for secondary schools in England to estimate the effect of school inputs on cognitive skills as measured by test scores at the end of compulsory schooling, at age 16, on a sample of 360,000 siblings attending same schools. We let cognitive skills at age 16 depend on cognitive skills observed at the end of primary school, at age 11, in the context of a value added model. Moreover, we let the return to school inputs vary across children with different levels of past skills, and we attribute the differential school expenditure effect to complementarity. While many previous papers on effects of school inputs have taken into account the dynamics of the educational production function by allowing past cognitive skills to affect present skills, they do not allow for complementarity between school inputs and past skills. Despite considerable interest in the technology of skill formation, we know of no previous empirical evidence on complementarity between school inputs and past skills. Presenting this evidence is the main contribution of this paper.

The challenge in making causal inference on complementarity is that both exogenous variation in child investments and exogenous variation in past abilities are needed. Past ability is endogenous in education production models because both past and current cognitive skills may depend on unobservable omitted characteristics such as cognitive and non-cognitive endowments and family inputs. We control for the endogeneity of past skills by adopting child fixed effects estimation, which also allows us to eliminate other unobserved confounding influences. The estimation of child fixed effects is possible because for each child we observe current and past test scores in the subjects of English, Mathematics and Science in our data. Moreover, we are able to exploit idiosyncratic variation in school expenditure caused by the fact that the allocation of funding to schools adjusts to actual school needs with delay and is heterogenous across areas, therefore causing variation in the expenditure within and across schools which is not explained by school characteristics, time dummies or school-specific trends.

We find robust evidence of both complementarity and self-productivity. The persistence in cognitive ability is 0.221 and the return to school expenditure is three times higher for students at the top of the past attainment distribution than for those at the bottom. A £1,000 increase in the expenditure per student increases test scores by 3% of a standard

deviation for children at the bottom of the past test score distribution, while it leads to an increase of 9% of a standard deviation in the test score for children at the top of the past test score distribution. So skills obtained at the end of primary school both persist into secondary school and raise the productivity of school inputs in secondary school.

We perform a number of robustness checks to make sure that these results are not caused by specialization in one subject, subject-specific parental investments, or unobserved abilities with differential effects across subjects. We also check the robustness of our results to changes in the definitions of school inputs and to the adoption of a natural metric for the test scores. Finally, we discuss the reliability of our assumptions on measurement error and on homogeneity of our model coefficients across subjects and expenditure levels.

In summary our empirical results suggest that the return to school inputs is higher for higher ability students. This confirms that there is an equity-efficiency trade-off for investments during secondary school (late child investments) as suggested by Cunha et al. (2006). The implications for policies that allocate funds to schools are therefore in line with those formulated by the child development literature, and suggest adjusting the balance of funding between elementary and secondary education in favor of the earlier years.

References

- Aizer A. and F. Cunha (2012), “The Production of Human Capital in Childhood: Endowments, Investments and Fertility”, National Bureau of Economic Research, NBER Working Paper Series, 18429.
- Almond D. and B. Mazumder (2013), “Fetal origins and parental responses”, *Annual Review of Economics*, 5, 37-56.
- Altonji J.G. and T.A. Dunn (1996), “Using Siblings to Estimate the Effect of School Quality on Wages”, *The Review of Economics and Statistics*, 78(4): 665-671.
- Amini, C. and S. Commander (2012), “Educational scores: how does Russia fare?”, *Journal of Comparative Economics*, 40, 3, 508-527.
- Andrabi T., Das J., Khwaja A.I., and T. Zajonc (2011), “Do Value-Added Estimates Add Value? Accounting for Learning Dynamics”, *American Economic Journal: Applied Economics*, 3(3): 29-54.
- Behrman J.R., M.R. Rosenzweig and P. Taubman, (1996), “College Choice and Wages: Estimates Using Data on Female Twins”, *The Review of Economics and Statistics*, 78(4): 672-685.
- Bond T.N., Lang K. (2013). “The Evolution of the Black-White Test Score Gap in Grade L-3: The Fragility of Results”, *The Review of Economics and Statistics*, 95(5), 1468-1479.
- Borghans, L., A. L. Duckworth, J. J. Heckman, and B. Ter Weel (2008), “The Economics and Psychology of Personality Traits”, *Journal of Human Resources*, 43, 972-1059.
- Carneiro, P. and J.J. Heckman (2002). “The evidence on credit constraints in post-secondary schooling”, *Economic Journal* 112 (482), 705-734.
- Carneiro, P. and J.J. Heckman, J (2003), “Human capital policy”, in Heckman, J.J., A.B. Krueger and B.M. Friedman (eds.), *Inequality in America: What Role for Human Capital Policies?* MIT Press, Cambridge, MA.
- Chowdry, H. and L. Sibieta (2011), “School Funding Reform: An Empirical Analysis of Options for a National Funding Formula”, IFS Briefing Note BN123.
- Clotfelter, C.T., H.F. Ladd and J.L. Vigdor (2010), “Teacher Credentials and Student Achievement in High School. A Cross-Subject Analysis with Student Fixed Effects”, *Journal of Human Resources* 45(3): 655-681.
- Cunha, F. and J.J. Heckman (2007), “The Technology of Skill Formation”, *American Economic Review*, 92(2): 31-47.
- Cunha F. and J.J. Heckman (2008), “Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation”, *Journal of Human Resources*, 43(4), 738-782.
- Cunha F., J.J. Heckman, L. Lochner L. and D. Masterov (2006), “Interpreting the Evidence on Life Cycle Skill Formation”, in Hanushek E. and Welch F. (eds.), *Handbook of the Economics of Education*, North Holland, 697-812.

- Cunha F., J.J. Heckman and S. Schennach (2010), “Estimating the Technology of Cognitive and Noncognitive Skill Formation”, *Econometrica*, 78 (3), 883-931.
- Cunha F. (2011), “Recent Developments in the Identification and Estimation of Production Functions of Skills”, *Fiscal Studies*, 32 (2), 297-316.
- Dee, T.S. (2005), “A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?”, *American Economic Review Papers and Proceedings* 95(2), 158-165.
- Dee, T.S. (2007), “Teachers and the Gender Gaps in Student Achievement”, *Journal of Human Resources* 42(3), 528-554.
- Del Boca, D., C. Monfardini and C. Nicoletti (2012), “Children’s and Parent’s Time-Use Choices and Cognitive Development during Adolescence” Human Capital and Economic Opportunity Working Group working paper 2012-006, Chicago: University of Chicago.
- Del Boca D., C. Flinn C. and M. Wiswall (2014), “Household Choices and Child Development”, *Review of Economic Studies*, forthcoming.
- Department for Education (2011), A Consultation on School Funding Reform: Rationale and Principles. <http://www.education.gov.uk/consultations/downloadableDocs/School%20Funding%20Reform%20consultation%20final.pdf>, accessed 18.1.2012
- Gibbons, S., and S. McNally (2013), “The Effects of Resources Across School Phases: A Summary of Recent Evidence”, CEP Discussion Paper No 1226.
- Eide, E., and M.H. Showalter (1998), The effect of school quality on student performance: A quantile regression approach, *Economics Letters*, Elsevier, 58, 3, 345-350.
- Figlio, D.N. (1999), “Functional form and the estimated effects of school resources”, *Economics of Education Review*, 18, 2, 241-252.
- Hanushek E.A. (1986), “The Economics of Schooling: Production and Efficiency in Public Schools” *Journal of Economic Literature*, 24, 1141-1177.
- Hanushek E.A. (1997), “Assessing the Effects of School Resources on Student Performance: An Update”, *Educational Evaluation and Policy Analysis*, 20, 19, 141-164.
- Hanushek, E. A. (2003), “The failure of input-based schooling policies”, *Economic Journal*, 113 (February), F64-98.
- Hanushek E.A., S.G. Rivkin and L.L. Taylor (1996), “Aggregation and the Estimated Effects of School Resources”, *The Review of Economics and Statistics* 78: 4, 611-627.
- Heckman, J.J., J. Stixrud, and S. Urzua (2006), “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior”, *Journal of Labor Economics*, 24, 411-482.
- Holmlund, H., S. McNally and M. Viarengo (2010), “Does money matter for schools?”, *Economics of Education Review*, 29, 1154-1164.
- Jenkins, S.P. and C. Schluter (2002), “The effect of family income during childhood on later-life attainment: evidence from Germany”, IZA Discussion Paper 604.

- Jepsen C. and S. Rivkin (2009), “Class Size Reduction and Student Achievement: The Potential Tradeoff between Teacher Quality and Class Size”, *Journal of Human Resources*, 44, 1, 223-250.
- Krueger, A. (1998), “Reassessing the view that american schools are broke”, *Economic Policy Review of the Federal Reserve Bank of New York*, 4 (1), 29-46.
- Levačič, R. (2008), “Financing Schools. Evolving Patterns of Autonomy and Control”, *Educational Management Administration & Leadership*, 36(2): 221-234.
- Levy, D. and G.J. Duncan (2000), “Using sibling samples to assess the effect of childhood family income on completed schooling”, Working paper, JCPR.
- Meghir C., and S. Rivkin (2011), “Econometric Methods for Research in Education”, in Hanushek E.A., S. Machin and L. Woessmann (eds.), *Handbook of the Economics of Education*, Volume 3, 1-87.
- Morris, P., G.J. Duncan and E. Clark-Kauffman (2005), “Child well-being in an era of welfare reform: The sensitivity of transitions in development to policy change”, *Developmental Psychology* 41 (6), 919-932.
- Mueller, S. (2013), “Teacher experience and the class size effect. Experimental evidence”, *Journal of Public Economics*, 98, 44-52.
- Nicoletti, C. and B. Rabe (2012), “The effect of school resources on test scores in England” ISER working paper 2012-13, Colchester: University of Essex.
- Rangvid B.S. (2007), “School composition effects in Denmark: quantile regression evidence from PISA 2000”, *Empirical Economics*, 33, 359-388.
- Rivkin S.G., Hanushek E.A., and J.F. Kain (2005), “Teachers, schools, and academic achievement”, *Econometrica*, 73(2), 417-458
- Rosenzweig, M. and K.I. Wolpin (1994), “Are there Increasing Returns to the Intergenerational Production of Human Capital? Maternal Schooling and Child Intellectual Achievement”, *Journal of Human Resources*, 29(2): 670-693.
- Summers, A. A., and B.L. Wolfe (1977). “Do Schools Make a Difference?”, *American Economic Review*, 67, 4, 639-652.
- Slater H., Davies N.M., and S. Burgess (2010), “Do Teachers Matter? Measuring the Variation in Teacher Effectiveness in England”, *Oxford Bulletin of Economics and Statistics*, 74 (5), 629-645.
- Spearman, C.E. (1904), “General intelligence, Objectively Determined And Measured”, *American Journal of Psychology*, 15: 201-293.
- Todd P. and K.I. Wolpin (2003), “On the Specification and Estimation of the Production Function for Cognitive Achievement”, *Economic Journal*, 113, 485, F3-F33.
- Todd, P.E. and K.I. Wolpin (2007), “The Production of Cognitive Achievement in Children: Home, School and Racial Test Score Gaps”, *Journal of Human Capital*, 1(1), 91-136.

West, A. (2009), "Redistribution and Financing Schools in England under Labour. Are Resources Going Where Needs Are Greatest?", *Educational Management Administration & Leadership* 37(2): 158-179.

Tables

Table 1: Difference in school expenditure between younger and older sibling

	No. of sibling pairs	Sibling difference in expenditure	
		mean	std. dev.
Siblings at same school			
<i>gross</i>	179,735	£349	£283
<i>net of time trend</i>	179,735	£36	£254
<i>net of school-specific time trend</i>	179,735	£15	£371
Siblings at different schools			
<i>gross</i>	31,982	£268	£754
<i>net of time trend</i>	31,982	-£48	£748
<i>net of school-specific time trend</i>	31,982	-£66	£376

Notes: National Pupil Database, 2007-2010; Consistent Financial Reporting Data 2005-2010; Schools, Pupils and their Characteristics Data 2005-2010. Pupil expenditure in 2010 prices, calculated using GDP deflator. The sibling difference is the 3-year average school expenditure of the younger sibling minus 3-year average school expenditure of the older sibling.

Table 2: Descriptive statistics: Individual and school-level controls

	mean	std. deviation
<i>Individual characteristics</i>		
Male	0.503	
No. school-age siblings in state schools	2.572	0.838
Older sibling or singleton in observation window	0.500	
First language English	0.918	
White British	0.865	
Black	0.019	
Mixed	0.022	
Indian	0.023	
Pakistani/Bangladeshi	0.043	
Chinese	0.003	
Other ethnicity	0.024	
Free school meal eligible	0.104	
Gifted and talented	0.184	
Special Educational Need, with statement	0.013	
Special Educational Need, no statement	0.150	
Deprivation score of residence	0.188	0.162
No. months older than August-born	5.490	3.478
<i>School characteristics (3 year averages)</i>		
Expenditure per student (£/1000)	4,905	0.666
Number of pupils (full time equivalent)	1,169	349
Prop. free school meal eligible	0.122	0.102
Prop. first language English	0.908	0.169
Prop. Special Educational Need, with statement	0.021	0.012
Prop. Special Educational Need, no statement	0.162	0.083
Prop. white	0.851	0.204
Prop. black	0.027	0.063
Prop. mixed	0.025	0.023
Prop. Indian	0.023	0.069
Prop. Pakistani/Bangladeshi	0.035	0.107
Prop. Chinese	0.003	0.005
Prop. other ethnicity	0.032	0.043
KS2 English scores, by cohort	26.9	1.3
KS2 Maths scores, by cohort	27.4	1.5
KS2 Science scores, by cohort	28.9	1.2
Number of observations	359,470	

Notes: National Pupil Database, 2007-2010; Schools, Pupils and their Characteristics Data 2005-2010.

Table 3: The effect of school expenditure per student: main results

	(1)	(2)	(3)
	OLS	Sibling fixed effects	Benchmark: two-step sib FE
Heterogenous effect of expenditure per student			
at the top decile	0.029** (0.007)	0.085** (0.007)	0.094** (0.005)
Difference to the top decile for			
9th decile	-0.013** (0.001)	-0.009** (0.001)	-0.011** (0.001)
8th decile	-0.022** (0.001)	-0.017** (0.001)	-0.020** (0.001)
7th decile	-0.025** (0.001)	-0.019** (0.001)	-0.023** (0.001)
6th decile	-0.031** (0.001)	-0.025** (0.001)	-0.030** (0.001)
5th decile	-0.037** (0.002)	-0.030** (0.001)	-0.034** (0.001)
4th decile	-0.041** (0.002)	-0.035** (0.001)	-0.040** (0.001)
3rd decile	-0.046** (0.002)	-0.041** (0.001)	-0.045** (0.002)
2nd decile	-0.049** (0.003)	-0.047** (0.002)	-0.051** (0.002)
1st decile	-0.040** (0.003)	-0.052** (0.002)	-0.060** (0.002)
Persistence ρ	0.521** (0.005)	0.378** (0.003)	0.221** (0.004)
F, H0: diff to top decile=0	157.32	110.17	103.47
	(4)	(5)	(6)
	OLS	Sibling fixed effects	Benchmark: two-step sib FE
Homogenous effect of expenditure per student			
at the mean	-0.001 (0.006)	0.055** (0.007)	0.061** (0.007)
Persistence ρ	0.578** (0.002)	0.452** (0.001)	0.304** (0.001)
Observations	1,078,410	1,078,410	1,078,410

Notes: + $p < .10$, * $p < .05$, ** $p < .01$. Student expenditure in 2010 prices, calculated using GDP deflator. Tests are standardized separately by subject. Standard errors clustered at school level. Control variables include all variables listed in Table 2 plus the standardized lagged test and dummies for academic year. Pooled sample, pooling the observations for Mathematics, English and Science (sample size is therefore three times the number of observations reported in Table 2).

Data source: National Pupil Database, 2007-2010; Consistent Financial Reporting Data 2005-2010; Schools, Pupils and their Characteristics Data 2005-2010.

Table 4: Difference in the effect of expenditure per student by gender

	Boys	Girls
	Two-step	Two-step
	sibling fixed effects	sibling fixed effects
Effect of expenditure per student at the top decile	0.116** (0.005)	0.078** (0.005)
Difference to the top decile for		
9th decile	-0.015** (0.001)	-0.007** (0.001)
8th decile	-0.024** (0.001)	-0.015** (0.001)
7th decile	-0.029** (0.001)	-0.017** (0.001)
6th decile	-0.037** (0.001)	-0.022** (0.001)
5th decile	-0.043** (0.002)	-0.025** (0.002)
4th decile	-0.050** (0.002)	-0.028** (0.002)
3rd decile	-0.058** (0.002)	-0.031** (0.002)
2nd decile	-0.066** (0.002)	-0.036** (0.003)
1st decile	-0.074** (0.003)	-0.044** (0.003)
Persistence ρ	0.219** (0.005)	0.221** (0.005)
F, H0: diff to top decile=0	97.20	30.10
Observations	542,511	535,899

Notes: + $p < .10$, * $p < .05$, ** $p < .01$. Student expenditure in 2010 prices, calculated using GDP deflator. Tests are standardized separately by subject. Standard errors clustered at school level. Control variables include all variables listed in Table 2 plus the standardized lagged test and dummies for academic year. Pooled sample, pooling the observations for Mathematics, English and Science.

Data source: National Pupil Database, 2007-2010; Consistent Financial Reporting Data 2005-2010; Schools, Pupils and their Characteristics Data 2005-2010.

Table 5: Robustness checks: subject-specific endogeneity

	(1)	(2)	(3)
	Benchmark model	Excl. groups with 'learning anomalies'	Science & Maths only
Effect of expenditure per student at the top decile	0.094** (0.005)	0.103** (0.004)	0.088** (0.006)
Difference to top decile for			
9th decile	-0.011** (0.001)	-0.007** (0.001)	-0.003** (0.001)
8th decile	-0.020** (0.001)	-0.013** (0.001)	-0.008** (0.001)
7th decile	-0.023** (0.001)	-0.013** (0.001)	-0.007** (0.001)
6th decile	-0.030** (0.001)	-0.018** (0.001)	-0.014** (0.001)
5th decile	-0.034** (0.001)	-0.019** (0.002)	-0.017** (0.002)
4th decile	-0.040** (0.001)	-0.022** (0.002)	-0.022** (0.002)
3rd decile	-0.045** (0.002)	-0.023** (0.002)	-0.028** (0.002)
2nd decile	-0.051** (0.002)	-0.024** (0.003)	-0.038** (0.003)
1st decile	-0.060** (0.002)	-0.026** (0.003)	-0.052** (0.003)
Persistence ρ	0.221** (0.004)	0.272** (0.005)	0.148** (0.005)
F, H0: diff to top decile=0	103.47	21.2	54.8
Observations	1,078,410	648,819	718,940

Notes: + $p < .10$, * $p < .05$, ** $p < .01$. Student expenditure in 2010 prices, calculated using GDP deflator. Tests are standardized separately by subject. Standard errors clustered at school level. Control variables include all variables listed in Table 2 plus the standardized lagged test and dummies for academic year. Pooled sample, pooling the observations for Mathematics, English and Science.

Data source: National Pupil Database, 2007-2010; Consistent Financial Reporting Data 2005-2010; Schools, Pupils and their Characteristics Data 2005-2010.

Table 6: Robustness checks: Alternative measures of expenditure and test scores

	(1)	(2)	(3)		(4)
	Benchmark model	“Real” outcome: Prob. of post-16 schooling partic.	“Real” input measure: teacher expenditure teachers	other	Expenditure in exam year only
Effect of expenditure per student at the top decile	0.094** (0.005)	1.899** (0.142)	0.122** (0.009)	0.047** (0.006)	0.061** (0.003)
Difference to top decile for					
9th decile	-0.011** (0.001)	0.060** (0.012)	-0.019** (0.004)	0.000 (0.006)	-0.011** (0.001)
8th decile	-0.020** (0.001)	0.056** (0.017)	-0.037** (0.004)	0.003 (0.006)	-0.019** (0.001)
7th decile	-0.023** (0.001)	0.077** (0.022)	-0.048** (0.005)	0.011 (0.007)	-0.022** (0.001)
6th decile	-0.030** (0.001)	-0.007 (0.028)	-0.068** (0.005)	0.020** (0.007)	-0.029** (0.001)
5th decile	-0.034** (0.001)	-0.046 (0.035)	-0.073** (0.005)	0.017* (0.007)	-0.033** (0.001)
4th decile	-0.040** (0.001)	-0.144** (0.042)	-0.088** (0.006)	0.023** (0.008)	-0.038** (0.001)
3rd decile	-0.045** (0.002)	-0.252** (0.051)	-0.099** (0.006)	0.025** (0.008)	-0.043** (0.001)
2nd decile	-0.051** (0.002)	-0.413** (0.062)	-0.112** (0.007)	0.026** (0.009)	-0.049** (0.002)
1st decile	-0.060** (0.002)	-0.654** (0.075)	-0.131** (0.009)	0.029* (0.011)	-0.057** (0.002)
Persistence ρ	0.221** (0.004)	0.390** (0.006)		0.214** (0.004)	0.222** (0.004)
F, H0: diff to top decile=0	103.47	34.45	32.95	2.34	101.85
Observations	1,078,410	1,078,410	1,078,410		1,078,410

Notes: + $p < .10$, * $p < .05$, ** $p < .01$. Student expenditure in 2010 prices, calculated using GDP deflator. Tests are standardized. Robust standard errors in parenthesis. Control variables include all variables listed in Table 2 plus the standardized lagged test and dummies for academic year. Pooled sample, pooling the observations for Mathematics, English and Science. Column (4) uses as transformation of current and lagged test scores the predicted probability of participating in post-compulsory schooling.

Data source: National Pupil Database, 2007-2012; Consistent Financial Reporting Data 2005-2010; Schools, Pupils and their Characteristics Data 2005-2010.

Table 7: Robustness checks: heterogeneity by subject and level of expenditure

	(1)			(2)	
	Heterogeneity by subject			Heterogeneity by level of expenditure	
	Maths	English	Science	low	high
Effect of expenditure per student at the top decile	0.034** (0.006)	0.052** (0.006)	0.167** (0.006)	0.097** (0.005)	0.097** (0.005)
Difference to top decile for					
9th decile	-0.009** (0.001)	-0.011** (0.001)	-0.006** (0.001)	-0.011** (0.001)	-0.012** (0.001)
8th decile	-0.011** (0.001)	-0.021** (0.001)	-0.014** (0.001)	-0.020** (0.001)	-0.020** (0.001)
7th decile	-0.011** (0.001)	-0.027** (0.001)	-0.017** (0.001)	-0.025** (0.001)	-0.024** (0.001)
6th decile	-0.015** (0.001)	-0.029** (0.001)	-0.023** (0.001)	-0.032** (0.001)	-0.031** (0.001)
5th decile	-0.018** (0.001)	-0.033** (0.001)	-0.024** (0.001)	-0.036** (0.002)	-0.036** (0.001)
4th decile	-0.020** (0.001)	-0.036** (0.001)	-0.032** (0.001)	-0.042** (0.002)	-0.042** (0.002)
3rd decile	-0.028** (0.002)	-0.037** (0.002)	-0.033** (0.002)	-0.047** (0.002)	-0.048** (0.002)
2nd decile	-0.037** (0.002)	-0.035** (0.002)	-0.038** (0.002)	-0.054** (0.003)	-0.056** (0.002)
1st decile	-0.057** (0.002)	-0.028** (0.003)	-0.038** (0.003)	-0.062** (0.004)	-0.067** (0.003)
Persistence ρ	0.271** (0.004)	0.260** (0.004)	0.221** (0.004)	0.235** (0.005)	0.193** (0.005)
F, H0: diff to top decile=0	86.63	94.53	61.12	48.65	67.67
Observations		1,078,410		1,078,410	

Notes: + $p < .10$, * $p < .05$, ** $p < .01$. Student expenditure in 2010 prices, calculated using GDP deflator. Tests are standardized separately by subject. Standard errors clustered at school level. Control variables include all variables listed in Table 2 plus the standardized lagged test and dummies for academic year. Pooled sample, pooling the observations for Mathematics, English and Science

Data source: National Pupil Database, 2007-2010; Consistent Financial Reporting Data 2005-2010; Schools, Pupils and their Characteristics Data 2005-2010.

A Appendix: asymptotic bias of value-added model

Because $Y_{iht,11}$ and $u_{iht,16}$ in model (4) are correlated we would generally expect the ordinary least squares estimator to be biased and inconsistent; but, under the above assumption that Y_{iht}^* and $e_{iht,t}$ have equal persistence, the asymptotic bias caused by this correlation cancels out. Indeed we can prove that the asymptotic bias of the ordinary least square estimation of ρ is equal to:

$$plim \hat{\rho}_{FEE} = \rho + \left[\frac{Cov(\mu_{ih}, M_W Y_{iht,11})}{Var(M_W Y_{iht,11})} \right] + \left[\frac{Cov(e_{iht,16}, e_{iht,11})}{Var(M_W Y_{iht,11})} - \rho \frac{Var(e_{iht,11})}{Var(M_W Y_{iht,11})} \right], \quad (11)$$

where $W = [I_{ih}^F, I_{ih}^S, X_{ih}]$ is the vector of explanatory variables in our value added model (4), which excludes the lagged test and the unobserved child specific endowment, and M_W is the projection matrix on the space orthogonal to the one generated by the variables W . The first term between brackets is the asymptotic bias caused by omission of the child endowment μ_{ih} ; while the second term between square brackets is the asymptotic bias caused by the correlation between $Y_{iht,11}$ and $u_{iht,16}$, which cancels out because the assumption of identical persistence in Y_{iht}^* and in $e_{iht,t}$ implies that

$$Cov(e_{iht,16}, e_{iht,11}) = \rho Var(e_{iht,11}).$$

B Appendix: estimation and consistency of heterogeneous model

Let us consider the education production model with heterogeneous effects of school investment

$$Y_{iht,16} = \alpha + I_{ih}^F \beta_F + I_{ih}^S \beta_{S,10} + \sum_{j=1}^9 I_{ih}^S \mathbf{I}(d_{j-1}^* < Y_{iht,11}^* \leq d_j^*) \beta_{S,j} + X_{ih} \gamma + Y_{iht,11}^* \rho + \mu_{ih} + e_{iht,16}. \quad (12)$$

As for the model with homogenous effect we assume that

$$Y_{iht,11} = Y_{iht,11}^* + e_{iht,11} \quad \text{and} \quad Y_{iht,16} = Y_{iht,16}^* + e_{iht,16}. \quad (13)$$

We assume also that each of the three subject-specific ability $Y_{ihs,16}$ and the latent ability $Y_{ih,16}^*$ follow the same dynamics, which implies that

$$e_{ihs,16} = \rho e_{ihs,11} + \sum_{j=1}^9 I_{ih}^S [\mathbf{I}(d_{j-1}^* < Y_{ihs,11} \leq d_j^*) - \mathbf{I}(d_{s,j-1} < Y_{ihs,11} \leq d_{s,j})] \beta_{S,j} + v_{ihs,16}. \quad (14)$$

where $d_{s,j}$ is the j -th decile of $Y_{ihs,11}$ and, as before, $v_{ihs,16}$ is identically and independently distributed across children, households and test subjects with mean zero, homoscedastic and independent of the inputs in the production model and of the true latent skill at age 11 and 16, $Y_{ih,11}^*$ and $Y_{ih,16}^*$.

To estimate model (12) we replace the unobserved latent ability $Y_{ih,11}^*$ with the lagged test score in subject s , $Y_{ihs,11}$, and the unobserved $\mathbf{I}(d_{j-1}^* < Y_{ih,11}^* \leq d_j^*)$ with $\mathbf{I}(d_{s,j-1} < Y_{ihs,11} \leq d_{s,j})$, and we consider the following model

$$Y_{ihs,16} = \alpha + I_{ih}^F \beta_F + I_{ih}^S \beta_{S,10} + \sum_{j=1}^9 I_{ih}^S \mathbf{I}(d_{s,j-1} < Y_{ihs,11} \leq d_{s,j}) \beta_{S,j} + X_{ih} \gamma + Y_{ihs,11} \rho + \mu_{ih} + u_{ihs,16}, \quad (15)$$

where $u_{ihs,16} = e_{ihs,16} - \sum_{j=1}^9 I_{ih}^S [\mathbf{I}(d_{j-1}^* < Y_{ihs,11} \leq d_j^*) - \mathbf{I}(d_{s,j-1} < Y_{ihs,11} \leq d_{s,j})] \beta_{S,j} - \rho e_{ihs,11}$, which, using (14), simplifies to $u_{ihs,16} = v_{ihs,16}$.

We can estimate model (15) by using a two-step estimation procedure. In the first step we consider the child fixed effect estimation of the regression

$$Y_{ihs,16} = \rho Y_{ihs,11} + \sum_{j=1}^9 I_{ih}^S \mathbf{I}(d_{s,j-1} < Y_{ihs,11} \leq d_{s,j}) \beta_{S,j} + \mu_{ih} + v_{ihs,16}. \quad (16)$$

In the second step we use the estimated coefficients ρ and $\beta_{S,j}$ to estimate

$$Y_{ihs,16} - \rho Y_{ihs,11} - \sum_{j=1}^9 I_{ih}^S \mathbf{I}(d_{j-1} < Y_{ihs,11} \leq d_j) \beta_{S,j} = \alpha + I_{ih}^F \beta_F + I_{ih}^S \beta_{S,10} + X_{ih} \gamma + \mu_{ih} + u_{ihs,16}, \quad (17)$$

using sibling fixed effects to control for unobserved family heterogeneity.