THE UNIVERSITY *of York*

Discussion Papers in Economics

No. 13/30

School inputs and skills: Complementarity and self-productivity

Cheti Nicoletti & Birgitta Rabe

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

School inputs and skills: Complementarity and self-productivity

Cheti Nicoletti

DERS, University of York and ISER, University of Essex

Birgitta Rabe

ISER, University of Essex

This version: November 20, 2013

Abstract

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 0.22 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.

Keywords: Education production function, test scores, school quality, complementarity
JEL codes: I22, I24

Contact: Cheti Nicoletti (cheti.nicoletti@york.ac.uk), Birgitta Rabe (brabe@essex.ac.uk)

We thank the Department for Education for making available, under Terms and Conditions, data from the National Pupil Database. We thank Sonia Bhalotra, Stephanie von Hinke Kessler Scholder and participants of the European Economic Association Meeting 2013 in Gothenburg for helpful comments and suggestions. Financial support from the Nuffield Foundation is gratefully acknowledged. The Nuffield Foundation is an endowed charitable trust that aims to improve social well-being in the widest sense. It funds research and innovation in education and social policy and also works to build capacity in education, science and social science research. The Nuffield Foundation has funded this project, but the views expressed are those of the authors and not necessarily those of the Foundation. More information is available at www.nuffieldfoundation.org. Part of the work was also supported by the Economic and Social Research Council through the Research Centre on Micro-Social Change, award no. RES-518-28-001.

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 interest in the technology of skill formation, solid empirical evidence of complementarity that could shed light on this is difficult to come by. 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 0.22 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, focusing on children from disadvantaged backgrounds. We are able to take account of the redistributive allocation of school resources by controlling for the school characteristics that 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 to be 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). Our inference on complementarity is free of this criticism because it is based on child fixed effect estimation that controls for differences in unobserved child skills, be they cognitive, non-cognitive or other skills. By adopting child fixed effects estimation we are also able to eliminate other confounding influences, such as peer spillover effects that are not subject specific.

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 identification strategy used to produce our empirical evidence on self-productivity and complementarity. Section 3 describes the data, while Section 4 presents the main results and robustness checks. Finally, Section 5 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:

$$Y^*_{ih,16} = f(I^F_{ih}, I^S_{ih}, X_{ih}, Y^*_{ih,11}, \mu_{ih}), \tag{1}$$

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^F_{ih} is the family investment in child cognitive development between ages 11 and 16, I^S_{ih} 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 $\frac{\partial f(.)}{\partial Y^*_{ih,11}} > 0$. Furthermore, we consider two different specifications of model (1), one where the productivity of school inputs does not change across level of the 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, $\frac{\partial^2 f(.)}{\partial I^S_{ih} \partial Y^*_{ih,11}} > 0$.

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 is an extension of 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 model[1]

$$Y_{ihs,11} = Y^*_{ih,11} + e_{ihs,11} \text{ and } Y_{ihs,16} = Y^*_{ih,16} + e_{ihs,16}, \tag{2}$$

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

[1]Imposing 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). In Nicoletti and Rabe (2012) we also report the results of two factor analyses of the three test scores at ages 11 and 16. In both analyses we find that the first factor explains more than 80% of the total variance and its factor loadings are all about 0.9.

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. $Cov(e_{ihs,16}, e_{ihs,11}) \neq 0$.[2] Furthermore, we assume that the persistence in $Y_{ih,t}$ be 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 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) be additive, separable and linear in its arguments, we can rewrite it as

$$Y_{ihs,16} = \alpha + I^F_{ih}\beta_F + I^S_{ih}\beta_S + X_{ih}\gamma + Y^*_{ih,11}\rho + \mu_{ih} + e_{ihs,16}, \tag{3}$$

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_I 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

$$Y_{ihs,16} = \alpha + I^F_{ih}\beta_F + I^S_{ih}\beta_S + X_{ih}\gamma + Y_{ihs,11}\rho + \mu_{ih} + u_{ihs,16}, \tag{4}$$

[2]Note 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 attitude 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.

where $u_{ihs,16} = e_{ihs,16} - e_{ihs,11}\rho$. Because $Y_{ihs,11}$ and $u_{ihs,16}$ are correlated we would generally expect the ordinary least squares estimator to be biased and inconsistent; but, under the above assumption that $Y^*_{ih,t}$ and $e_{ihs,t}$ 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:

$$plim\ \hat{\rho}_{FFE} = \rho + \left[\frac{Cov(\mu_{ih}, M_W Y_{ihs,11})}{Var(M_W Y_{ihs,11})}\right] + \left[\frac{Cov(e_{ihs,16}, e_{ihs,11})}{Var(M_W Y_{ihs,11})} - \rho\frac{Var(e_{ihs,11})}{Var(M_W Y_{ihs,11})}\right], \quad (5)$$

where $W = [I^F_{ih}, I^S_{ih}, X_{ih}]$ 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 $Y_{ihs,11}$ and $u_{ihs,16}$, which cancels out because the assumption of identical persistence in $Y^*_{ih,t}$ and in $e_{ihs,t}$ implies that

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

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 the three contemporaneous tests and the three corresponding lagged tests 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 remaining slope coefficients because the corresponding variables do not vary across the three tests.

In the second step 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,

$$Y_{ihs,16} - Y_{ihs,11}\rho = \alpha + I^F_{ih}\beta_F + I^S_{ih}\beta_S + X_{ih}\gamma + \mu_{ih} + u_{ihs,16}. \quad (6)$$

For this regression we adopt *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^F_{ih} 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). Under the assumption that the inputs in the educational production function may depend on unobserved child-specific endowment only through past test scores and that the difference between siblings in the unobserved family characteristics is either zero or uncorrelated with the sibling difference in school investments once we control for sibling differences in the lagged test and in the other control variables, the sibling fixed effect estimation is consistent. We discuss the credibility of these assumptions in more detail in Section 2.3.

2.2 Specification with heterogenous effect of school investment

Model (3) imposes a marginal effect of school investment that is constant across different levels of lagged ability, i.e.

$$\frac{\partial Y^*_{ih,16}}{\partial I^S_{ih}} = \beta_S. \tag{7}$$

To allow for a heterogenous effect of school investment across different levels of lagged ability, we can modify model (3) by adding among the explanatory variables an interaction term between school investment and lagged cognitive ability, $(I^S_{ih} Y^*_{ih,11})$,

$$Y_{ihs,16} = \alpha + I^F_{ih}\beta_F + I^S_{ih}\beta_S + (I^S_{ih}Y^*_{ih,11})\beta_{SY} + X_{ih}\gamma + Y^*_{ih,11}\rho + \mu_{ih} + e_{ihs,16}, \tag{8}$$

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

$$\frac{\partial Y^*_{ih,16}}{\partial I^S_{ih}} = \beta_S + Y^*_{hs,11}\beta_{SY}. \tag{9}$$

Notice however that model (8) 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

$$\frac{\partial Y^*_{ih,16}}{\partial I^S_{ih}} = \beta_{S,10} + \mathbf{I}(d^*_{j-1} < Y^*_{ih,11} \leq d^*_j)\beta_{S,j}, \tag{10}$$

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,...,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,

$$Y_{ihs,16} = \alpha + I^F_{ih}\beta_F + I^S_{ih}\beta_{S,10} + \sum_{j=1}^{9} I^S_{ih}\mathbf{I}(d^*_{j-1} < Y^*_{ih,11} \leq d^*_j)\beta_{S,j} + X_{ih}\gamma + Y^*_{ih,11}\rho + \mu_{ih} + e_{ihs,16}. \quad (11)$$

To estimate model (11) 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 $j = 1, ..., 9$. 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 (11) and on the assumptions imposed for its consistency in Appendix A.

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 $\beta_{S,j}$ ($j = 1, ..., 9$) in model (11) even in the presence of unobserved student specific capabilities that 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 neither differ nor have a differential effect across subjects.[3] In other words, 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 (11).

While our inference on the presence of self-productivity and complementarity is consistent even in the presence of 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 (11). 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 special educational needs, i.e. learning difficulties that are related to behavioral or health issues, and

[3] Research for the UK indicates that more students at secondary school level receive private paid tutoring in Mathematics than in English or Science (see Ireson 2004), but this may be because parents are better able to help in non-Mathematics subjects at home. To check for potential differences between humanistic and scientific subjects we replicate our estimation considering only Mathematics and Science in our robustness checks (see Section 4.3 for the results).

for being identified by the school as gifted and talented. This implies that our estimation allows for sibling differences in parental investments that depend on sibling differences in special educational needs, in talent as well as in lagged cognitive ability.[4]

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. 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 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).

[4]Another 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.

[5]In 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.

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 net of a time trend (we control for academic year in our models). The Table shows that the mean difference in expenditure per student net of the 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 time trend. 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 and Further Education with student characteristics and learning aims.[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.

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

[6] The 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.

[7] We 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.

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.

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

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

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.

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 an indicator of whether the school is a community school[9] or not and the number of students in the school. We also characterize schools in terms of their student composition, using 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).

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

[9] Community schools are owned, governed and managed by the Local Education Authority, whereas in voluntary aided and voluntary controlled schools as well as in foundation schools some or all of these functions are carried out by other organizations such as the Church of England in faith schools, for example.

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

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

[10] Missing 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%

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 invariant across 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 heterogenous 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 the test score 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 monotonically increasing in the lagged test score and the differences between the students in two adjacent deciles are always statistically significantly different form 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. The only

[11]We do not bootstrap the standard errors to take account of the fact that we replace ρ and $\beta_{S,j}$ for $j = 1, ..., 9$ 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.

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.

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 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.

Neglecting unobserved family and school characteristics and/or the endogeneity of the lagged test (columns 1 and 2) leads to an underestimation of the effect of expenditure per student and an overestimation of the persistence in the cognitive ability in both the models with homogenous and heterogenous effect of school expenditure (see bottom and top panels of Table 3). Since the omitted variables (family and school characteristics or unobserved child endowment in the case of neglected endogeneity) 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 equation (5)). 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 first 9 deciles of the lagged test, $\mathbf{I}(d^*_{j-1} < Y^*_{ih,11} \leq d^*_j)$ for $j = 1, ..., 9$, at least when we expect the effect of school inputs to be highest for student at the top decile of the lagged test distribution, i.e. if $\beta_{S,j} < 0$ for $j = 1, ..., 9$, and if the omitted interaction terms are negatively correlated with the lagged test after controlling for the remaining control variables in the production model.

To check for potential differences by gender in the learning process, we also estimate the education production model separately by gender and we report in Table 4 the results of our preferred estimation, i.e. the two-step estimation of the value added model that allows for a heterogenous effect of the expenditure per student by level of lagged cognitive ability. 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 estimations in the second step. Overall, the separate results for boys and girls show that complementarity also holds for cuts of the data: the effect of school expenditure increases monotonically with lagged test scores. The results are quite interesting; while the persistence in the cognitive ability is almost identical for boys and girls at 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.

4.2 Restricted value added models

We have seen in the last section that the omission of unobserved family and school characteristics and the endogeneity of the lagged test cause an underestimation of the effect of the expenditure per student and an overestimation of the persistence. But what happens if we impose perfect or no persistence in test scores, i.e. if we consider a restricted value added model assuming that $\rho = 1$ or $\rho = 0$? The assumption of perfect persistence ($\rho = 1$) is standard in the so-called restricted value-added or gain score model (see Summers and Wolfe 1977; Hanushek 2003; Rivkin et al. 2005 for examples). The assumption of zero persistence ($\rho = 0$) is sometimes taken because of data limitations (see Krueger 1998; Todd and Wolpin 2003; Clotfelter et al. 2010 for examples) and is equivalent to considering a model where the test score is explained only by current inputs. This model is referred to as contemporaneous model (see Todd and Wolpin 2003 for details on the assumptions of this model).

Imposing $\rho = 1$ in our value added model with homogenous effect is equivalent to considering a new education production model where the dependent variable is the difference between the test scores at age 16 and at age 11 and the explanatory variables are given by current inputs. This can be estimated using sibling fixed effects estimation rather than our two-step estimation. The results show that adopting the homogenous model with $\rho = 1$ leads to an underestimation of the school expenditure effect, which almost halves compared to the unrestricted model (compare column 1 and 2 in the bottom panel of Table 5). This underestimation is in line with our theoretical expectation because an overestimation of ρ leads to an underestimation of the school expenditure effect (see also Andrabi et al. 2011).

Our two-step estimation of the restricted value added model that imposes perfect persistence and allows heterogenous returns to school inputs produces an effect of 0.036 at the 5th decile, which is very close to the effect found without allowing for heterogeneity, 0.035 (compare column 2 in the top and bottom panels of Table 5). However, the return to school inputs varies dramatically from the top to the bottom decile, with very negative effects of school expenditure for children at the top deciles of the lagged test distribution and very positive effects for children at the bottom deciles. This is because when we impose perfect persistence in the test scores, we wrongly assume that student attainment at the end of secondary school is as good as it was at the end of primary school. If this perfect persistence does not hold, then imposing $\rho = 1$ implies erroneously predicting too high (low) test scores at the end of compulsory schooling for high (low) achievers at the end of primary school. The estimation of the school expenditure effect at different deciles in the restricted model with $\rho = 1$ counterbalances this bias by producing a large negative (large positive) effect of school expenditure for children at the top (bottom) deciles. In other words, the decreasing rather than increasing effect of expenditure per student when moving from the bottom to the top of the lagged test distribution is the result of wrongly imposing $\rho = 1$.

Imposing $\rho = 0$ in our value added model with homogenous effect assumes either that past inputs do not contribute to the production of skills or that inputs are invariant across time and capture the effect of both contemporaneous and past inputs. Both of these assumptions are restrictive and we expect that the contemporaneous model will lead to an overestimation of the effect of current inputs compared to the unrestricted value added model. To show this we report in Table 5 the results for the restricted value added model that imposes zero persistence in test scores. This can be estimated using sibling fixed effects estimation with the age 16 test scores as outcome. The estimated effect of expenditure per student, when we assume a homogenous effect, is approximately 12% higher than the estimated effect of the corresponding value added model with no restriction on ρ. When comparing the model with heterogenous effects and unrestricted ρ and the corresponding model imposing zero persistence, we find that the effect of expenditure per student is largely over-estimated at any decile of the lagged test except the first three deciles, where the effect is underestimated (compare columns 1 and 3 of Table 5). Notice that, as expected, imposing $\rho = 0$ has a similar but oppositive effect than imposing $\rho = 1$, i.e. it causes an overestimation of the strength of complementarity between school inputs and lagged abilities.

In conclusion, restricted education production models that impose perfect or zero persistence can lead to considerably different estimates of the effect of school spending than unrestricted models if this effect is assumed to be homogenous across pupils with different levels of past skills. If we allow the return to school inputs to vary across students (heterogenous effect), the results are extremely different and can seriously bias inference on the dynamics of cognitive skill production.

4.3 Robustness checks

Our preferred estimation indicates that the effect of expenditure per student is increasing across the distribution of the 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 (i) strong preferences for specific subjects, (ii) the existence of more than one type of cognitive ability, (iii) the selection into our sample.

Preferences and 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, 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 girls (see Table 4). But this does not seem to be the case because we find that the average difference between the highest test score across Mathematics, Science and English 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 average standardized test score in Mathematics between age 11 and 16 is smaller for boys than for girls.

Our second concern is the possibility that the test scores in Mathematics, Science and English are measuring different types of cognitive abilities. Mathematics and Science are subjects for which more analytic skills (fluid intelligence) might be required than for English. To assess whether this can be an issue we consider a factor analysis of the three test scores in Mathematics, English and Science and check whether there is more than one latent factor explaining the variation in these three test scores. We find that the first factor explains more than 80% of the total variance and the factor loadings of this first factor are equal to 0.9 for all three test scores (see Nicoletti and Rabe 2012 for details on these results). For this reason we conclude that the assumption that test scores in Mathematics, Science and English measure the same type of cognitive ability is acceptable. To check the hypothesis that test scores in Mathematics and Science measure more analytic abilities that are different from the ability measured by English test scores, we also compute the two-step estimation excluding English test scores. Table 6, column (2) 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. This check also assures us that unobserved parental investments do not have a differential effect across subjects. It may be possible, for example, that parents are better able to help their children in English than in analytic subjects such as Mathematics and Science, but even if this was the case it does not seem to affect the estimation of the effect of expenditure per student.

We also want to consider whether the persistence in test scores is equal across the three subjects we consider, as this is an assumption of our model. To test this empirically we allow the coefficient ρ in the first step of our two-step estimation procedure to have a different value for each subject. Considering our model with a homogenous effect of school inputs, we find that the persistence of Science and English are very similar, 0.28 and 0.30. The persistence of the Mathematics test score is only slightly higher at 0.34, but this could be because measurement error on Mathematics tests is smaller than on English and Science tests. In Nicoletti and Rabe (2012) we show that measurement error bias is a minor concern in value added models.

Regarding possible sample selection effects, we consider the results of the two-step estimation when excluding London. This is because London is often seen as a special case with higher teacher wages and staff turnover and younger teachers. The results in column (3) of Table 6 show that the estimates do not vary from our preferred estimates. Finally, we also consider results when including in the estimation sample both siblings going to the same and different schools. Here we find that the effect of expenditure is much reduced, and this is likely the cause of a correlation between the sibling difference in school expenditure and in their unobserved endowments. This confirms that we prefer the sample of same school siblings only.

5 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 the same school. 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 and time dummies.

We find robust evidence of both complementarity and self-productivity. The persistence in cognitive ability is 0.22 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 or by sample selection, and that they do not change by choosing a different set of subjects.

We compare our education production model with a heterogenous effect of school inputs with models commonly used in the education literature. Imposing the same effect of the return to school inputs across students with different levels of lagged cognitive ability (homogenous model) leads to an overestimation of the persistence in skills. The homogenous value added model overestimates the effect of an increase of school spending for children with low abilities and underestimate the effect for children with high abilities. Similar biases but of much larger magnitude are found when wrongly imposing a perfect persistence in test scores. On the contrary, if we wrongly impose a zero persistence the strength of the complementarity is overestimated. We also find that neglecting to control for the endogeneity of past cognitive ability leads to an overestimation of the persistence of cognitive ability and to an underestimation of the strength of complementarity.

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-effciency 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.

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), 705734.

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.

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.

Ireson, J. (2004), "Private Tutoring: how prevalent and effective is it?", *London Review of Education* 2(2): 109-122.

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 Research 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), 919932.

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", *The 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			
gross	179,735	£349	£283
net of time trend	179,735	£36	£254
Siblings at different schools			
gross	31,982	£268	£754
net of time trend	31,982	-£48	£748

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 on the younger sibling minus 3-year average school expenditure on the older sibling.

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

	mean	std. deviation
Individual characteristics		
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
School characteristics (3 year averages)		
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
Community school	0.587	
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) OLS	(2) Sibling fixed effects	(3) Two-step sibling fixed effects
Heterogenous effect of expenditure per student			
At the top decile	0.029**	0.085**	0.094**
	(0.007)	(0.007)	(0.005)
Difference to the top decile for			
9th decile	-0.013**	-0.009**	-0.011**
	(0.001)	(0.001)	(0.001)
8th decile	-0.022**	-0.017**	-0.020**
	(0.001)	(0.001)	(0.001)
7th decile	-0.025**	-0.019**	-0.023**
	(0.001)	(0.001)	(0.001)
6th decile	-0.031**	-0.025**	-0.030**
	(0.001)	(0.001)	(0.001)
5th decile	-0.037**	-0.030**	-0.034**
	(0.002)	(0.001)	(0.001)
4th decile	-0.041**	-0.035**	-0.040**
	(0.002)	(0.001)	(0.001)
3rd decile	-0.046**	-0.041**	-0.045**
	(0.002)	(0.001)	(0.002)
2nd decile	-0.049**	-0.047**	-0.051**
	(0.003)	(0.002)	(0.002)
1st decile	-0.040**	-0.052**	-0.060**
	(0.003)	(0.002)	(0.002)
Persistence ρ	0.521**	0.378**	0.221**
	(0.005)	(0.003)	(0.004)
	(4) OLS	(5) Sibling fixed effects	(6) Two-step sibling fixed effects
Homogenous effect of expenditure per student			
At the mean	-0.001	0.055**	0.061**
	(0.006)	(0.007)	(0.007)
Persistence ρ	0.578**	0.452**	0.304**
	(0.002)	(0.001)	(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. 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 (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 Two-step sibling fixed effects	Girls Two-step sibling fixed effects
Effect of expenditure per student		
at the top decile	0.116**	0.078**
	(0.005)	(0.005)
Difference to the top decile for		
9th decile	-0.015**	-0.007**
	(0.001)	(0.001)
8th decile	-0.024**	-0.015**
	(0.001)	(0.001)
7th decile	-0.029**	-0.017**
	(0.001)	(0.001)
6th decile	-0.037**	-0.022**
	(0.001)	(0.001)
5th decile	-0.043**	-0.025**
	(0.002)	(0.002)
4th decile	-0.050**	-0.028**
	(0.002)	(0.002)
3rd decile	-0.058**	-0.031**
	(0.002)	(0.002)
2nd decile	-0.066**	-0.036**
	(0.002)	(0.003)
1st decile	-0.074**	-0.044**
	(0.003)	(0.003)
Persistence ρ	0.219**	0.221**
	(0.005)	(0.005)
Observations	542,511	535,899

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

Table 5: The effect of expenditure per student when imposing $\rho = 1$ and $\rho = 0$.

	Unrestricted value added free ρ	Restricted value added $\rho = 1$	Contemporaneous model $\rho = 0$
Heterogenous effect of expenditure per student			
At the top decile	0.094**	-0.139**	0.160**
	(0.005)	(0.005)	(0.005)
Difference to the top decile for			
9th decile	-0.011**	0.044**	-0.026**
	(0.001)	(0.000)	(0.000)
8th decile	-0.020**	0.075**	-0.046**
	(0.001)	(0.001)	(0.001)
7th decile	-0.023**	0.109**	-0.061**
	(0.001)	(0.001)	(0.001)
6th decile	-0.030**	0.140**	-0.078**
	(0.001)	(0.001)	(0.001)
5th decile	-0.034**	0.175**	-0.093**
	(0.001)	(0.001)	(0.001)
4th decile	-0.040**	0.210**	-0.110**
	(0.001)	(0.001)	(0.001)
3rd decile	-0.045**	0.250**	-0.129**
	(0.002)	(0.001)	(0.001)
2nd decile	-0.051**	0.308**	-0.153**
	(0.002)	(0.001)	(0.001)
1st decile	-0.060**	0.419**	-0.195**
	(0.002)	(0.001)	(0.001)
Persistence ρ	0.221**	1	0
	(0.001)		
Homogenous effect of expenditure per student			
At the mean	0.061**	0.035**	0.072**
	(0.007)	(0.007)	(0.008)
Persistence ρ	0.304**	1	0
	(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. 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.
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: comparing different samples

	(1) Base sample: siblings same school two-step sibFE	(2) Science & Maths only two-step sibFE	(3) excluding London two-step sibFE	(4) incl. siblings at different schools two-step sibFE
Effect of expenditure per student				
at the top decile	0.094**	0.088**	0.099**	0.063**
	(0.005)	(0.006)	(0.005)	(0.003)
Difference to the top decile for				
9th decile	-0.011**	-0.003**	-0.011**	-0.011**
	(0.001)	(0.001)	(0.001)	(0.000)
8th decile	-0.020**	-0.008**	-0.020**	-0.020**
	(0.001)	(0.001)	(0.001)	(0.001)
7th decile	-0.023**	-0.007**	-0.024**	-0.023**
	(0.001)	(0.001)	(0.001)	(0.001)
6th decile	-0.030**	-0.014**	-0.031**	-0.030**
	(0.001)	(0.001)	(0.001)	(0.001)
5th decile	-0.034**	-0.017**	-0.035**	-0.035**
	(0.001)	(0.002)	(0.001)	(0.001)
4th decile	-0.040**	-0.022**	-0.041**	-0.041**
	(0.001)	(0.002)	(0.001)	(0.001)
3rd decile	-0.045**	-0.028**	-0.047**	-0.046**
	(0.002)	(0.002)	(0.002)	(0.001)
2nd decile	-0.051**	-0.038**	-0.054**	-0.052**
	(0.002)	(0.003)	(0.002)	(0.002)
1st decile	-0.060**	-0.052**	-0.063**	-0.061**
	(0.002)	(0.003)	(0.002)	(0.002)
Persistence ρ	0.221**	0.148**	0.220**	0.217**
	(0.004)	(0.005)	(0.004)	(0.003)
Observations	1,078,410	718,940	978,006	1,270,302

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

A Appendix

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

$$Y_{ihs,16} = \alpha + I^F_{ih}\beta_F + I^S_{ih}\beta_{S,10} + \sum_{j=1}^{9} I^S_{ih}\mathbf{I}(d^*_{j-1} < Y^*_{ih,11} \leq d^*_j)\beta_{S,j} + X_{ih}\gamma + Y^*_{ih,11}\rho + \mu_{ih} + e_{ihs,16}. \quad (12)$$

As for the model with homogenous effect we assume that

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

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

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

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

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

where $u_{ihs,16} = e_{ihs,16} - \sum_{j=1}^{9} I^S_{ih}[\mathbf{I}(d^*_{j-1} < Y_{ihs,11} \leq d^*_j) - \mathbf{I}(d_{s,j-1} < Y_{ihs,11} \leq d_{s,j})]\beta_{S,j} - \rho e_{ihs,11}$, 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

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

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

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

using sibling fixed effects to control for unobserved family heterogeneity.