What Makes a Test Score ? The Respective Contributions of Pupils, Schools and Peers in Achievement in English Primary Education*

Francis Kramarz† Stephen Machin‡ Amine Ouazad §

December 2007

Abstract

The aim of this study is to determine how pupil ability and background, peers, and school quality affect educational achievement. To do so we use rich data from England that matches pupils with their peers and schools. The dataset records all English pupils and their test scores in Key Stage 1 (age 7) and Key Stage 2 (age 11) examinations. The quality of the data source, coupled with our econometric techniques, allows us to assess the respective importance of different educational inputs. In addition, we can distinguish school effects, that affect all pupils irrespective of their year and grade of study, from school-grade-year effects. Identification of the pupil effects separately from these school-grade-year effects is achieved because students are mobile across schools. Peer effects are identified assuming variations in school-grade-year group composition in adjacent years are exogenous. We estimate three different specifications, the most complete allowing Key Stage 2 results to be affected by the Key Stage 1 school(-grade-year) at which the pupil studied. We discuss the validity of our various exogeneity assumptions. Estimation results suggest the following ranking of inputs: pupils' ability and background are more important than school time-invariant inputs. Peer effects are significant but small.

*We would like to thank audiences of various seminars and conferences. In particular, Marc Gurgand, Eric Maurin, Thomas Piketty suggested helpful improvements. Previous versions of this paper were presented at the May 2006 COST meeting, the Jamboree of the European Doctoral Program, the Lunch Seminar of Paris-Jourdan Sciences Economiques, the Crest internal seminar, the 2006 IZA prize in labor economics conference, the summer research program at Boston College and the networks seminar at Cornell University. The authors acknowledge financial and computing support from the Centre for Economic Performance, the Centre for the Economics of Education, CREST-INSEE and the Economics of Education and Education Policy in Europe (EEEPE) program. Programs are available from the corresponding author, Amine Ouazad.

†CREST, CEPR and IZA

‡University College London, Centre for Economic Performance, London School of Economics and CEPR

§Centre for Economic Performance at the London School of Economics, Paris School of Economics (PSE) and CREST; Centre for Economic Performance, Houghton Street, London WC2A 2AE, United Kingdom, a.ouazad@lse.ac.uk

1 Introduction

A key question for education policy is which of the many educational inputs - including social background, schools, peers and teachers - really make a difference? The answer is crucial to good decision-making in education. Educational policies in a context of limited funding face a trade-off between a range of policies like targeting pupils, targeting schools, promoting desegregation, implementing tracking, hiring and promoting good teachers. Choosing between these alternative strategies requires some knowledge on the relative impact of schools, pupils' abilities, family background and peers on educational achievement.

If schools make a difference then changing school inputs, management and teaching practices could enhance educational performance and may lead to substantial reductions of inequalities. If, on the other hand, peers are more important, segregation may be the number one issue to tackle. Finally, if pupils' ability or background - more generally pupil-specific issues - are the principal determinants of achievement, policies targeting low achieving pupils may have the highest potential to narrow the gap between children.

However, there is no real consensus on what really *makes a difference* (Summers & Wolfe 1977). This is partly because the questions of how to identify these different effects are very challenging from an empirical viewpoint. This said, there are some stylised facts that emerge from different strands of the literature. For example, scholars in the sociology of education have long argued that, apart from students' ability and background, peers are the most important determinant of test scores. This dates back at least as far as the Coleman (1966). In *The Concept of Equality of Educational Opportunity* (1969), Coleman asserts that:

> [...] those inputs characteristics of schools that are most alike for Negroes and whites have least effect on their achievement. The magnitudes of differences between schools attended by Negroes and those attented by whites were as follows: least, facilities and curriculum; next, teacher quality; and greatest, educational backgrounds of fellow students. The order of importance of these inputs on the achievement of Negro students is precisely the same: facilities and curriculum least, teacher quality next, and backgrounds of fellow students, most.

Following the Coleman report, a series of desegregation programs were initiated – notably bussing programs. Furthermore the report sparked research on the effects of peers, school quality and pupil's backgrounds on achievement.

But a number of papers by economists have challenged this viewpoint. A seminal paper (Manski 1993) has highlighted the main problems of the baseline specification used by James Coleman. Selection bias is the most important one: if we observe good pupils together, are they good because they are together or are they together because they are good? Students may be partly selected on unobservable characteristics. Moreover, Manski (Manski 2000, Manski 1993) pointed out that it is hard to disentangle the effect of peers' behaviour from the effect of peers' characteristics. And last, the econometrician should address the issue of simultaneity bias since students influence each other simultaneously. Hoxby (2000) has estimated the overall effect of race and gender composition on Texas primary school pupils. She finds significant and large peer-effects. In the context of the Boston METCO desegregation program, Angrist & Lang (2002) estimate the effect of minority students on test scores. Their estimated effects are modest and short-lived. Gould, Lavy & Paserman (2004) assess the impact of immigrants on Israeli pupils. Even though the average effect they find are not significant, they do find that low-achieving pupils are more sensitive to their peers.

Another strand of the literature has focused on the relationship between school quality and achievement. Usually school quality has been proxied by the teacher-pupil ratio, teacher education, teacher experience, teacher salary or expenditures per pupil. Overall, the link between school resources and test scores is weak (Hanushek 1986, Hanushek 2003). The 'school effectiveness' research, carried out mostly by educationalists, comes to a similar conclusion: schools matter, but not by anywhere near as much as non-school factors like the home environment (Mortimore, Sammons, Stoll, Lewis & Ecob 1988, Stiefel, Schwartz, Rubenstein & Zabel 2005, Teddlie & Reynolds 1999, West & Pennell 2003). In the British context, Levacic & Vignoles (2002) mention that the impact of school resources is small and very sensitive to misspecification. Dearden, Ferri & Meghir (2002) suggest that, while the pupil-teacher ratio has no significant impact, attending selective schools improves both attainment and wages. Eventually the literature turns to less crude measures of school quality. Indeed,

> Schools differ dramatically in quality, but not for the rudimentary factors that many researchers (and policy makers) have looked to for explanation of these differences. (Hanushek 1986)

In a recent paper, Rivkin, Hanushek & Kain (2005) indeed argue that test scores are the sum of student, school and teacher effects, potentially unobservable. Analysis should not therefore solely rely on observable characteristics for the estimation of school and teacher effectiveness. However, they do not try to identify all the different components of their favored specification.

The focus of our paper is rather different to the work that focusses on a single issue. Rather we attempt to measure the relative contributions of pupils, schools and peers without restricting our analysis to observable proxies for peers' characteristics or school quality. To do so we jointly estimate time-varying school fixed effects (school-grade-year effects) and pupil fixed effects. The former are then decomposed into an observable time-varying part (the social composition) and an unobservable school effect. Our estimation strategy combines ideas from the literature using matched worker-firm data Abowd, Kramarz & Margolis (1999) and from work in the economics of education, e.g. Hoxby (2000). Following Abowd et al. (1999), pupil and school effects are identified using movers, assuming in particular that mobility decisions are not motivated by time-varying pupil-specific shocks. Following Hoxby (2000), we argue that variations in the average quality of pupils within a school across consecutive years are essentially idiosyncratic, because demographics change randomly from year to year around a central tendency.

The paper goes further than the previous literature at least two reasons: (i) it assesses the relative contribution of peers, school quality and pupils' ability and background using a single equation; (ii) it estimates the overall effect of peers without relying on specific peer characteristics.

To estimate the various effects, we rely on an administrative database of English pupils. We use three cohorts of English pupils in state schools. The dataset is exhaustive; all such pupils are followed from primary education to secondary education. The outcomes measured are national test scores (Key Stage 1 and Key Stage 2). The grades achieved at the end of these Key Stages are particularly important instruments for both parents and the English education authorities. In particular, government uses them to set targets and parents can freely read them in performance tables, published on the web or in the popular press.

Based on various specifications, in terms of contributions to explain pupil performance, our estimation results show that pupil heterogeneity are a more important determinant of achievement inequalities than school quality. Peer effects are mostly small, significant for the impact of free school meal pupils, boys, Chinese pupils, Indian and Black Carribean pupils. We assess the robustness of our assumptions and examine the mobility patterns in the data. They largely confirm that conditioning on person effects and on school-grade-year effects is a reasonable strategy.

The finding that pupil effects matter most is, of course, important in the light of research arguing that early interventions (often pre-school) yield higher educational achievement returns (Heckman & Masterov 2007). If such policies aimed at dampening down achievement gaps on entry (or early on) in primary school do indeed work best then our findings suggest that this is likely to have an important impact on subsequent gaps and inequalities in educational achievement that occur throughout the compulsory school years. But our findings also show there to be important, albeit smaller, contributions of peers and schools to the variance of pupil achievement. This also matters for the design of education policies during the school careers of children.

The outline of the rest of the paper is as follows: Section 2 presents the various specifications that will be estimated as well as the estimation strategy for each such specification. In each case, the specification is related to papers of the literature. Section 3 introduces the reader to the specific British policy context and describes the dataset. Section 4 analyzes the regression results. Section 5 discusses the robustness of the estimation, examines the mobility patterns, and the public policy implications of the results.

2 The Econometric Model

2.1 Specifications

In this section we present the econometric models from which we intend to extract estimates of the relative importance of different educational inputs. The plan is to implement this model in the context of primary school children in England. The compulsory school careers of English children are organised into four Key Stages, Key Stages 1 and 2 which take place in primary school, and Key Stages 3 and 4 in secondary school. Our focus is on primary schools where Key Stage 1 examinations are taken at age 7 (grade 2) and Key Stage 2 examinations at the end of primary school at age 11 (grade 6)[1].

To begin note that, as Rivkin et al. (2005) point out, academic achievement at any point in a pupil's education is a cumulative function of her endowment (ability and family background), of school quality, and of the environment (community, in particular). This implies that: (i) test scores are a function of these educational inputs; (ii) these inputs can vary over time; and (iii) educational production functions should include the whole history of inputs that shaped each pupil's experience. The current section presents four different specifications that incorporate some or all of these features, as depicted below:

[1]Ideally, we would like to estimate the model up to key stage 4, but the dataset does not follow any cohort from key stage 1 to key stage 4. Future releases of the National Pupil Database will follow cohorts from key stage 1 to key stage 3 and above.

$$y_{i,f,t} = \mathbf{x_{i,f,t}}\beta + \theta_i + \psi_{J(i,t)} + \varepsilon_{i,f,t} \quad (1)$$

$$y_{i,f,t} = \mathbf{x_{i,f,t}}\beta + \theta_i + \varphi_{J(i,t),g(i,t),t} + \varepsilon_{i,f,t} \quad (2)$$

$$y_{i,f,t} = \mathbf{x_{i,f,t}}\beta + \theta_i + \varphi_{J(i,t),g(i,t),t} + \lambda\varphi_{J(i,t-1),g(i,t-1),t-1} + \varepsilon_{i,f,t} \quad (3)$$

And the last specification,

$$y_{i,f,t} = \mathbf{x_{i,f,t}}\beta + (1 + \lambda(t-1)) \cdot \theta_i + \varphi_{J(i,t),g(i,t),t} + \lambda\varphi_{J(i,t-1),g(i,t-1),t-1} + \varepsilon_{i,f,t} \quad (4)$$

In all these specifications, there are two Key Stage periods $t = 1, 2$, i denotes the N pupils with $i = 1, \ldots, N$; j denotes the J schools with $j = 1, \ldots, J$. $y_{i,f,t}$ is the test score of pupil i at time t in examination topic f. $J(i,t)$ denotes the school pupil i attended at time t and $g(i,t)$ denotes the grade in which pupil i attends in year t.

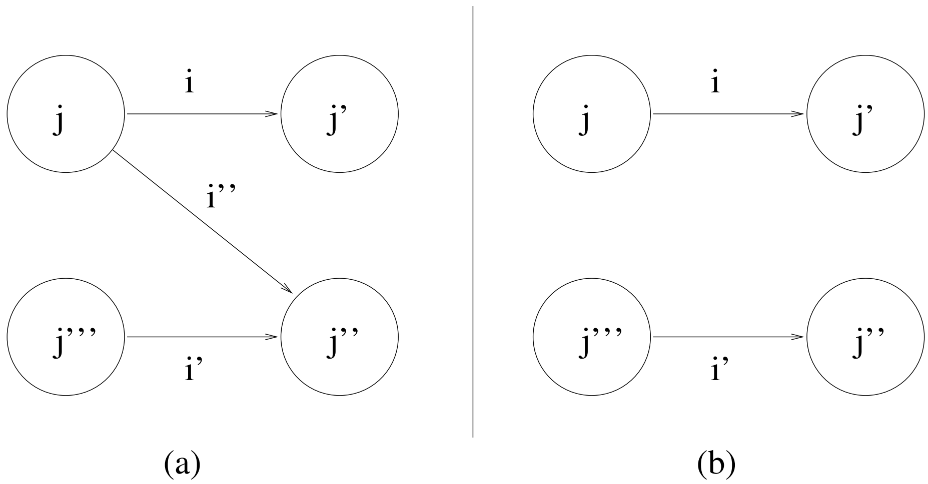
In the first specification, exactly similar to that chosen in Abowd et al. (1999), the test score is decomposed into a pupil effect θ_i, a school effect $\psi_{J(i,t)}$, $\mathbf{x_{i,f,t}}\beta$ is the effect of the K time-varying covariates, and $\varepsilon_{i,t}$ the residual.[2] The covariates contain controls for cohorts, years and topics.

The main advantage of specification (1) is its simplicity. However, it does not take into account the fact that school-specific inputs can vary over time. For instance, teachers may be different from one year to the other and the student body of the school changes. To take into account this feature, specification (2) therefore posits that achievement is the sum of a student effect and a school-grade-year effect.

But these first two specifications are still restrictive in that they assume the only input that affects outcomes at different stages of the educational curriculum is pupil ability. So, for example, the quality of the teachers or the environment at the initial stage does not affect future outcomes. It is likely however that some features of the school (or school-grade-year) that a pupil attended in the past (Key Stage 1, date $t-1$) have an impact on test scores at date t. This feature is captured in Specification (3) which states that test scores are the sum of a student effect, the current school-grade-year effect and the past school-grade-year effect discounted by λ. Because we do not observe grades before Key Stage 1, we constrain the initial past school-grade-year effect to be equal to zero, $\varphi_{j,g,0} = 0$. At this initial date, the school-grade-year effect and the pupil effect θ cannot be separately identified given the data. Notice that, in the same fashion as in value-added models, this specification constrains the current and past effect of schools to be proportional (Todd & Wolpin 2003).

Finally, there is one last issue remaining in specification (3), namely the child's progress only depends on the school and not on his ability. The most general specification (4) allows progress of the child to depend both on the child's ability and his past and current schools. It also allows us to assess the long run effect of schools on achievement. If λ is nonzero, schools have an effect not only on current achievement but also on achievement in the next period (i.e. grade 6 at the end of Key Stage 2 in our study context).

[2] Of course in Abowd et al. the effects were firms (our schools) and workers (our pupils).



Notations: i, i', i'' are pupils. j, j', j'', j''' are schools or school-grade-years. Reading: i connects school j and school j' because i attends school j in the first period, and i attends school j' in the second period.

Figure 1: (a) Sufficient mobility – The mobility graph has only one connex component . (b) Mobility is not sufficient – The mobility graph has two connex components, $\{j, j'\}$ and $\{j'', j'''\}$.

2.2 Identification Hypotheses for Specifications 1 to 4

The identification of specifications (1) to (4) requires both *sufficient mobility* and *exogeneous mobility*, both of which are defined in this sub-section, in addition to the traditional exogeneity of the other covariates. In addition to this, we will assume that at least one of our specifications is correct. For instance, if the true model involves a pure match effect, namely an unobserved component specific to both the pupil and the school because, say, some schools are better suited to more able pupils, then our estimates would have no clear meaning. We need to rule this possibility out.[3]

Mobility is sufficient when the mobility graph for pupils and schools is connected (Abowd et al. 1999, Abowd, Creecy & Kramarz 2002). The mobility graph is defined as follows: two schools are connected if and only if at least one pupil has attended both schools in different years. The mobility graph is connected when it has only one connex component. This is illustrated in Figure 1.

Moreover, exogeneity assumptions specific to our model with pupil and school effects are required (Abowd et al. 1999, Abowd et al. 2002). Or, formally, in the simplest specification, $E(\varepsilon_{i,f,t}|i, J(i,t) = j) = 0$; conditional on pupil's identity and identity of the school, the residual should have mean zero. For instance, if unmeasured unemployment shocks affect some pupils' parents and have an effect (e.g. through reduced income) on outcomes, then our estimates of the pupil effects might be biased.

Furthermore, pupils' mobility should also be exogenous with respect to grades. To understand this point, consider the set of pupils who attend school j in period 1 and school j' in period 2. This mobility is clearly useful, not to say necessary, since these movements allow the identification of the relative effectiveness of school j with respect to school j'. Ignoring the effect of covariates for the sake of clarity, this gives:

[3] It is ruled out all the more enthusiastically since such a specification cannot be estimated with available techniques for a problem of this magnitude.

$$E[\Delta y_{i,f}|i, J(i,2) = j', J(i,1) = j] = \psi_{j'} - \psi_j + E[\Delta\varepsilon_{i,f}|i, J(i,2) = j', J(i,1) = j] \quad (5)$$

with $\Delta y_{i,f} = y_{i,f,2} - y_{i,f,1}$ and $\Delta\varepsilon_{i,f} = \varepsilon_{i,f,2} - \varepsilon_{i,f,1}$. The effectiveness of school j can be identified separately from the effectiveness of school j' as soon as school j and school j' are connected. If ψ_j is identified, then $\psi_{j'}$ is identified. Setting for instance ψ_0 to zero and θ_0 to zero identifies all school and pupil effects. A formal definition of the mobility graph and the identification of the model are detailed in appendix A.

It needs to be noted, however, that identification through sufficient mobility is not enough. A reformulation of the exogeneity assumption is helpful in order to understand when the estimated effects are truly identified (unbiased). Taking the difference between grades at t and at $t-1$ yields $E[\Delta\varepsilon_{i,f}|i, J(i,2) = j', J(i,1) = j] = 0$, which holds when, in addition to the absence of potential unobserved shocks that directly affect grades of some of the pupils (see above), there is no unobserved shock that affect mobility decisions of pupils and make some of them leave school j for school j' *conditional* on pupil's and school's identity. To summarize, we say that, in such a case, mobility is exogenous. For instance, mobility is exogenous when the probability of moving from school j to school j' is not correlated with unobserved time varying shocks, see (Hanushek & Rivkin 2003). If, for instance, those families who experience an unemployment shock between the two periods are more likely (i) to make their child move to a bad school and (ii) to experience lower test scores due to their parents' joblessness, then the difference between bad and good schools might be underestimated.

Introducing school-grade-year effects in specifications (2) and (3) not only allows more flexibility in the measurement of school effectiveness, it also introduces more pupil mobility, since pupils necessarily change year-group between the two periods. In addition, because school-grade-year effects time-vary, the exogeneity assumptions are weaker than before.

Model (2) is a particular case of (3) in which the discounting factor is set to 0. Again, identification conditions are identical.

2.3 The identification of Peer Effects

Once we have estimated school-grade-year effects, we would also like to disentangle the effect of the social composition of the school-grade-year from the other inputs under reasonable identifying conditions.

The identification of peer effects is challenging. The main issues are described in Manski (1993). First, students may be sorted partly based on unobservable characteristics – for instance, teachers and students may not be randomly matched. Second, students influence each other which means it is hard to disentangle the effect of one on the other; in other words, there is a simultaneity bias. And third, it is hard to identify the effect of peers' characteristics from the effect of peers' behaviour.

We assume, as in Hoxby (2000) and Gould et al. (2004), that the year-to-year variations in school-grade-year composition are exogenous, *conditional* on the school identity (put differently, conditional on the school effect) essentially because of the randomness of the demographics. For instance, plus minus one black carribean student in a given year is probably an idiosyncratic variation. We therefore regress the school-grade-year effect on a school identifier and the composition of the school-grade-year.

$$\varphi_{j,g,t} = \psi_j + \hat{E}(\mathbf{z}|j,g,t)\gamma + \nu_{j,g,t} \tag{6}$$

For each school-grade-year j, g, t, $\hat{E}(\mathbf{z}|j,g,t)$ denotes the vector of average student characteristics, for instance the fraction of boys, the fraction of blacks, the fraction of free school meal students.

This strategy captures most of the bias due to non-random sorting of students between schools, essentially assuming that there is no correlation between changes in school-grade-year composition and unobservable school inputs.

Formally, year-to-year variations in school-grade-year student composition should be exogenous conditional on the school-by-grade fixed effect. Variations in school-grade-year composition should not correlated with unobserved time-varying school characteristics.

$$E[\nu_{j,g,t=2} - \nu_{j,g,t=1}|\hat{E}(\mathbf{z}|j,g,t=2) - \hat{E}(\mathbf{z}|j,g,t=1)] = 0 \tag{7}$$

$\hat{E}(\mathbf{z}|j,g,t=2) - \hat{E}(\mathbf{z}|j,g,t=1)$ is the year-to-year variation in school-grade-year composition. $\nu_{j,g,t=2} - \nu_{j,g,t=1}$ represents unobserved time-varying shocks in school quality.

This hypothesis adresses the first identification issue, namely the selection bias. The simultaneity bias is adressed through the use of a common school-grade-year effect for all students. Students "share" the same public good, which includes teacher characteristics and peer-effects.

However we are not able to separately identify the effect of peers' characteristics and the effect of peers' behaviour. Thus the vector of social interactions γ captures both of these peer-effects. γ is the *reduced form* peer-effect.

2.4 Decomposing Inequalities

The models allow us to decompose inequalities of test scores and test score gaps into components attributed to schools, peers and pupils' ability and background.

Test scores are the sum of the pupil effect, the year-group effect and the past year-group effect.

$$\begin{aligned} y_2 &= \theta + \varphi_2 + \lambda\varphi_1 + \varepsilon_2 \\ y_1 &= \theta + \varphi_1 + \varepsilon_1 \end{aligned}$$

where indices have been dropped for the sake of clarity. Moreover, school-grade-year effects can be decomposed into a permanent school effect and the effect of school composition.

$$\varphi = \psi + \overline{z}\gamma + \nu$$

Inequalities of educational achievement can therefore be decomposed into inequalities of school quality, inequalities of pupil ability and background, and inequalities due to different social contexts, for instance stemming from varying patterns of segregation. In the first period:

$$Var(y_1) = Cov(y, \theta) + Cov(y, \varphi_1) + Var(\varepsilon_1)$$

The first term is the component due to pupils' differences in ability and background. It is the sum of the heterogeneity in pupil ability and background, and the matching between pupil ability and school-grade-year quality, i.e. $Cov(y, \theta) = Var(\theta) + Cov(\theta, \varphi)$. The same decompositon applies to school-grade-years, $Cov(y, \varphi) = Var(\varphi) + Cov(\theta, \varphi)$. Hence, school-grade-year quality can widen inequalities in test scores if (i) school-grade-years are heterogeneous (ii) good pupils – high-θ pupils – are matched with good school-grade-years. Matching good school-grade-years to low-θ pupils should reduce inequalities.

This match between good school-grade-years and low-θ pupils can take two routes: (i) by fostering desegregation, i.e. decreasing $Cov(\gamma\bar{z}, \theta)$ (ii) by matching high-ψ schools with low-θ pupils.

2.5 Other Identification Methods in the Literature

This section compares the identification strategy of this paper with the identification strategies that have been introduced in past literature.

We first compare our models to the value-added models that are used in most papers on the measurement of school effectiveness. A seminal paper (Rivkin et al. 2005) uses this model to emphasize the importance of teacher effects on academic achievement. Finally, we compare the identification strategy of our paper to hierarchical linear models that have flourished in the educational literature.

i) Value Added Models

In value added models, the outcome variable is the progress of the pupil rather than the absolute test score. This is basically a dynamic panel data model in which the coefficient on the lagged outcome is constrained to be equal to 1. In our most complete specification 4, we include the lagged fixed effects. The results of our estimations and of value added models are comparable.

A value added model would decompose the progress of the child into a child fixed effect, a school-grade-year effect and a residual.

$$\Delta y_{i,f,t} = \mathbf{x_{i,f,t}} \cdot \beta + \theta_i + \varphi_{J(i,t),g(i,t),t} + u_{i,f,t} \tag{8}$$

$\Delta y_{i,f,t} = y_{i,f,t} - y_{i,f,t-1}$ is the progress of the child between two subsequent periods. Other notations are as before. $\mathbf{x_{i,f,t}}$ is a vector of time-varying controls. There are two differences between this model and specification 4: (i) firstly, the value added model constrains the effect of past achievement on current achievement so that $\lambda = 1$ (ii) secondly, the error structure in the value-added model is such that time varying unobservables can have long-term consequences on achievement. Indeed, the value-added specification can be rewritten as:

$$y_{i,f,t} = \mathbf{x_{i,f,t}}\beta + \mathbf{x_{i,f,t-1}}\beta + 2 \cdot \theta_i + \varphi_{J(i,t),g(i,t),t} + \varphi_{J(i,t-1),g(i,t-1),t-1} + u_{i,f,t} + u_{i,f,t-1} \tag{9}$$

This model is equivalent to specification 4 when $\lambda = 1$ and $u_{i,f,t} + u_{i,f,t-1} = \varepsilon_{i,f,t}$. Thus value-added models are equivalent to our model with a decay rate of 1, up to the error structure. Point estimates should therefore be equal in both models. Our results suggest that teacher effects are robust to a range of decay rates λ. Thus estimated effects from the value added model and our full model should be similar.

ii) Teacher Effects: Rivkin et al. (2005)

Rivkin, Hanushek and Kain (Rivkin et al. 2005) specify an educational production function in which student value-added is decomposed into a student effect, a school effect and a teacher effect.

$$\Delta y_{i,f,t} = \theta_i + \psi_{J(i,t)} + \tau_{T(i,f,t)} + \varepsilon_{i,f,t} \tag{10}$$

$\Delta y_{i,f,t}$ is the gain in student achievement of student i, in field f in year t. This specification adds a teacher effect $\tau_{T(i,f,t)}$ where $T(i,f,t)$ is the teacher of student i in field f in year t. In their paper, Rivkin et al do not identify all the effects, but rather use this specification as a blueprint and then estimate bounds for the variance of the teacher effects.

Specification (10) is remarkable in a number of ways. First, it does not include school-grade-year effects. Second, it includes teacher effects, which is an addition to specification (1) of our paper. School effects $\psi_{J(i,t)}$ will not capture year-to-year variations of the student body, and the teacher effect $\tau_{T(i,f,t)}$ is likely to include both time-varying teacher quality and year-to-year variations of the student body that are correlated with changes in teacher quality.

iii) Metropolitan Area Fixed Effects: Hanushek & Rivkin (2003)

In Hanushek & Rivkin (2003), educational progress is decomposed into a family effect and a metropolitan area fixed effect.

$$\Delta y_{i,f,t} = \theta_i + \theta_{i,t} + MSA_{M(i,t)} + \varepsilon_{i,f,t} \tag{11}$$

$\Delta y_{i,f,t}$ is the gain in student achievement of student i in topic f in year t. $MSA_{M(i,t)}$ is a metropolitan area fixed effect, where $M(i,t)$ is the metropolitan area of student i in year t. There are 27 MSAs in Texas.

Three features of specification (11) are noticeable: first, student effects can vary over time; second, school effects nor school-grade-year effects are present; and last, metropolitan area fixed effects do not vary over time. These elements have the following consequences. Student time-varying fixed effects are likely to capture other time varying inputs such as the metropolitan area social composition. Metropolitan area fixed effects are likely to capture public good quality and average school quality in the area, but not year-to-year variations of the social composition of the area.

iv) Hierarchical Models for the Analysis of School Effects

Empirical work in education has used hierarchical linear models in several papers, notably in Raudenbush & Bryk (1986), Bryk & Raudenbush (1992), Goldstein (2002) or Rao & Sinharay (2006). These models identify individual effects, school effects and potentially teacher effects using only cross-sectional data. Despite its light data requirements, an important limitation of multilevel modelling is that they nest individual effects, school effects and other determinants of educational achievement. Our model does not, since it uses pupil mobility to identify the effects. Multilevel models are written as a set of specifications at multiple levels, e.g. a student level equation and a school level equation. Equations are then combined to lead to a single specification.

Another hindrance is that Raudenbush & Bryk (1986) and Bryk & Raudenbush (1992) specify *random* effects and not *fixed* effects. The identification of random effects require the assumption of strict exogeneity, which implies that, for instance, school effects are orthogonal to covariates and to other random and/or fixed effects (Wooldridge 2002, p257). It is likely that, for instance students with a high or a low effect go to particular schools, or that schools with a good intake are particular schools. Thus it is unlikely that the orthogonality between the effects and between effects and covariates is a realistic assumption.

2.6 Estimation method

The estimation of the model presented in this section cannot proceed in the same way as usual Ordinary Least Squares estimation techniques. The number of right-hand side variables is at least the sum $J-1+N+K$ of the number of school effects, pupil effects and the number of covariates [4]. Usual packages try to invert the matrix of covariates which is time consuming and numerically unstable. (Abowd et al. 2002) have therefore developed an iterative estimation technique to estimate the baseline fixed effects model of equation 1. However, specifications with past *and* current school or school-grade-year effects required new identification proofs and estimation techniques. The programs are freely available and can be obtained by email.

The estimation proceeds in the following way: first, it starts by computing the variance-covariance matrix of the model with dummies and covariates. This a large and sparse matrix. Second, starting with a first guess for fixed effects and coefficients – usually zero –, the procedure iterates by updating the approximate solution. The sequence of approximate solutions is obtained by the conjugate gradient algorithm, that converges to the true solution if and only if the variance covariance matrix is invertible; this requires that the mobility graph has one connex component and that the covariates are linearly independent. More about the conjugate gradient can be found in Dongarra, Duff, Sorensen & van der Vorst (1991). Details of the computation of the variance-covariance matrix are given in appendix B and these computation techniques have given birth to a set of programs developed by the authors and available on the website of the Cornell Institute for Social and Economic Research.

3 Dataset and Estimation Method

3.1 The English Educational Context

The English educational system currently combines market mechanisms (many of which were introduced in Education Act of 1988) with a centralized assessment of students' understanding of the curriculum (Machin & Vignoles 2005). Therefore it has the advantage of providing us with fairly different management and funding structures and, at the same time, national exam results for all students.

The assessment system features a National Curriculum which sets out a sequence of Key Stages through the years of compulsory schooling: Key Stage 1 from ages 5 to 7, Key Stage 2 from ages 7 to 11,

[4] One school or school-grade-year effect is set to zero, one pupil effect is set to zero, and we add a constant.

Key Stage 3 from ages 11 to 14 and Key Stage 4 from ages 14 to 16. At the end of each Key Stage, pupils are assessed in the core disciplines: Mathematics, English and Science (not for Key Stage 1). These tests are nationally set and anonymously marked by external graders.

The English primary schooling system is also characterised by a variety of different management structures and funding sources. Community schools and voluntarily controlled schools, which cater for more than half of the student body, are controlled by the Local Education Authority (LEA), of which there are 150 in England. In the case of community and voluntarily controlled schools, the LEA owns the buildings and employs the staff. On the other hand, in voluntary aided and foundation schools, teachers are employed by the school governing body and the LEA has no legal right to attend proceedings concerning the dismissal or appointment of staff. [5] Funding differs from school to school too. While most state schools are funded by the government, voluntary aided schools contribute around 10% of the total capital expenditure.[6] These management and funding differences are likely to generate various kinds of incentives, and thus different educational outcomes for pupils.

3.2 The National Pupil Database

The National Pupil Database (NPD) is a comprehensive administrative register of all English pupils in state schools. Data is collected by the Department for Education and Skills and it is mandatory for all state schools to provide accurate data on pupils, who are followed from year to year through a Pupil Matching Reference. Thus panel data can be built by stacking consecutive years of the National Pupil Database.

The dataset provides rich information on pupils' characteristics: gender, free school meal status, special educational needs, and the ethnicity group. Pupils who receive free meals are the 15 to 20% poorest pupils. The ethnicity variable of our sample encodes the main ethnic code: White, Black Caribbean, Other Black, Pakistani, African Black, Mixed background, Bangladeshi, Indian, Chinese, Other Background.[7] It also provides some information on school management, ie whether they are community schools, foundation schools, voluntary aided or voluntary controlled schools, and so forth.

Test scores in English, Maths and Science are also included – the latter field only for Key Stage 2. These tests are externally set and marked. We have standardized test scores to a mean of 50 and a standard deviation of 10 to make results comparable from one level to the other and from year to year.

The structure of the panel data we have built from the NPD is shown in table 1.

4 Baseline Results

We now come to the policy questions of the introduction: do schools matter? do peers matter? what management structures make schools better?

[5] Code of Practice on LEA Schools Relationships, DfES 2001

[6] Source: Eurybase, 2006.

[7] The encoding of ethnicity has changed in the period we consider. It was therefore necessary to recode it in a time consistent manner.

Key Stage	1		2
Age	7		11
Grade	2		6
Examination results in	Maths English		Maths English Science
School types	Infant First or lower Primary	Infant & Junior First & Middle Primary	Junior Middle Junior High
Cohorts	2000 1999 1998		2004 2003 2002

Table 1: Description of the Dataset

4.1 The Heterogeneity of School and Pupil Effects

Do schools matter? The question has been extensively discussed in the sociological and educational literature as well as in recent papers in the economics of education (Rivkin et al. 2005). The estimations of school-grade-year effects and individual effects specifications gives us a means to assess the effect of school quality, peer effects and pupil background on achievement.

The estimation of specifications 1 to 4 gives pupil fixed effects, school-grade-year effects and school effects. School-grade-year effects are not available for the simple school effects model (specification 1). Three remarkable stylized facts come out from the variance covariance tables (table 4 to 7): (i) first, pupils are much more heterogenous than either school or school-grade-years effects (ii) individual effects explain a much larger share of the variance of test scores than school or school-grade-year effects.

A key finding is that the variance of pupil fixed effects is much larger than the variance of school-grade-year effects and school effects. The standard deviation of pupil effects is between 3.8 and 4.8 times larger than the standard deviation of school effects. This suggests that pupils are more heterogeneous than schools. This is also true when comparing the standard deviation of pupil effects and the standard deviation of school-grade-year effects. Nevertheless, pupil fixed effects are less precisely estimated than school-grade-year effects or school effects. Indeed, at most five observations per child are available whereas on average 250 observations per school-grade-year are available. To address this potential issue, we look at the correlation between test scores and the school or school-grade-year effects. Individual effects are imprecisely estimated but this should actually lower the correlation between individual effects and test scores[8].

The correlation between test scores and individual effects is between 0.791 and 0.833, i.e. between 5 and 6.7 times larger than the correlation between school effects and test scores. Covariance tables (tables 6 and 7) confirm that the explanatory power of pupil effects is much larger than the explanatory power of school effects. This is no surprise given the high correlation, the low variance of school effects and the much higher pupil heterogeneity.

Thus results strongly suggest that pupil effects are a more important determinant of test scores than school effects. They are interpreted as all educational experiences before age 7, including parental background, childcare, kindergarten. Therefore the idea that pupil effects explain most of the variance of test scores is in line with Heckman & Masterov (2007), Currie (2001) and Garces, Thomas & Currie (2002).

4.2 Pupil Effects: Disentangling Individual Effects From Peer Effects and School Quality

Pupil effects can be correlated with a range of individual characteristics (like ethnicity, gender, free school meal status, special needs and the child's month of birth). This analysis differs from an analysis on raw test scores since, under maintained assumptions, it looks at pupil effects free of peer effects and free of the correlation between school quality and observable characteristics.

Pupil effects are well explained by observable characteristics, the R-Square of the regressions of the fixed effects is around 40% (table 8). Coefficients are remarkably robust to different specifications. Moreover, the results are in line with descriptive statistics on test scores. The pupil fixed effects of free school meal children are 40 to 41% of a standard deviation lower. Family disadvantage is important, with free school meal pupils being the 10 to 20% poorest children in England. Chinese pupils are the best performing pupils, with a fixed effect 16% of a standard deviation higher than white pupils. Interestingly, indian pupils have a lower fixed effect than white pupils (6.4 to 7.8% of a standard deviation lower), whereas the test scores of indian pupils are higher than the test scores of whites. This suggests that basic descriptive statistics do not disentangle the effect of ethnicity from the effect of peers and the effect of school quality. Finally, male pupils have a higher fixed effect, by 2.5 to 2.6% of a standard deviation. This is certainly due to the fact that regression were carried out by pooling all subjects together. Boys in primary school are better at mathematics and science, whereas girls are better at English. There are therefore around three observations for which boys are better – maths in grades 2 and 6, science in grade 6 – and 2 observations for which girls are better – English in grades 2 and 6.

Table 8 confirms that pupil effects are well explained by observable characteristics, even if it is not a causal interpretation. Coefficients are consistent with basic descriptive statistics, and at the margin this analysis allows to disentangle pure individual effects from the social context, i.e. peers and school quality.

4.3 Effects of Social Composition on School Quality

The effect of the gender, ethnic and social composition of the school on its quality can be identified under the assumption that year-to-year variations in school composition are exogenous. Two stylized facts

[8]This, of course, holds provided measurement error is classical.

emerge from the covariance tables and the regression of school-grade-year effects on school composition and school effects: (i) the covariance between school-grade-year effects and test scores is comparable to the covariance between school effects and test scores (ii) some of the peer effects are significant but small – effect of the fraction of boys, free meals and special needs. Overall, results suggest peer effects to be significant, but small.

Covariance tables 6 and 7 suggest that the covariance between test scores and school effects is very similar to the covariance between test scores and school-grade-year effects; 5.6 vs. 7.7 for the school-grade-year specification, 4.6 vs. 4.2 for the past and current school-grade-year specification, and 4.7 vs. 4.0 for the full-fledged specification. Since school-grade-year effects can be decomposed into the school effect and the effect of social composition, these figures imply that peer effects are likely to be less important than school quality.

We then estimate the effect of the fraction of different social groups on school-grade-year effects, separately for grade 2 and grade 6. In grade 2, in the full-fledged specification, increasing the fraction of male students by 10% makes school-grade-year effects fall by 0.4% of a standard deviation[9] (table 9, column 3). This effect is robust to different specifications. In grade 6, the effect of the fraction of boys is positive, i.e. increasing the fraction of boys by 10 percentage points increases test scores by 0.9% of a standard deviation (table 10, column 3). The difference between grade 2 and grade 6 gender composition effects is likely to be due to the fact that grade 2 exams are in English and Maths, whereas grade 6 exams are in English, Maths and Sciences. Boys are better than girls in both science and maths, but not better in English. This effect is robust to the inclusion of past school-grade-year effects and past individual effects in the baseline specification. Most papers in the literature find a negative effect of boys on achievement both in English and in Maths, e.g. Hoxby (2000).

The fraction of free meal children has a detrimental effect on school-grade-year effects in grade 6; increasing the fraction of free meal children by 10 percentage points increases school-grade-year effects by 0.7% of a standard deviation (table 10, column 3).

In grade 6, ethnic composition has an effect on school quality. Chinese and Indian children exert a positive contextual effect, black carribean children exert a negative contextual effect. The effects are large: increasing the fraction of chinese students by 10 percentage points increases fixed effects by 4.6% of a standard deviation (table 10, column 3). These results are in line with the intuition that being surrounded by high performing peers is good for your test scores; chinese children are at the top of the test score distribution, and black carribean children are at the lower tail.

Special needs pupils have a negative impact on achievement under the identification assumption. Indeed, increasing the fraction of special needs students by 10% decreases school-grade-year effects by 6% of a standard deviation. Other effects are not significant in grade 2. The contextual effect of special needs students includes both the direct effect of interacting with special needs students and the effect that goes through teachers' and principals' behaviour (Todd & Wolpin 2003). Section 5 will look more closely at the identification assumptions.

[9]The standard deviation of test scores is 10.

4.4 School Quality and Management Structures

Are some schools better than other schools? Are some management structures better than other management structures? Even if school quality explains a small share of the variance of test scores, there may be scope for improvement in the way schools are structured. English schools can have a variety of different organizational structures; as was said in the presentation of the English educational context, some schools hire and dismiss their staff while in other schools the staff is recruited and dismissed by the Local Education Authority (table 3). The regressions of tables 11 and 12 suggest that there are beneficial effects of local recruitment of staff coupled with external control of the school board.

In both grade 2 and grade 6, voluntary controlled schools perform worse than community schools; like community schools, they cannot locally manage their human resources and do not own their assets. The main difference with community schools is that they are mostly Church of England schools.

Table 12 shows that voluntary aided schools, who can hire and dismiss staff locally, have a higher school fixed effect in grade 6, by 3 to 4 % of a standard deviation. The effect is not significant in grade 2 (table 11). Other types of schools can hire locally, e.g. foundation schools. These schools do not have a significantly higher school effect in grade 2 and grade 6. But the board of foundation schools is controlled by the Local Education Authority, whereas the board of voluntary controlled schools is mostly controlled by the foundation. Broadly speaking, schools with a high fixed effect recruit locally and the majority of their board is externally controlled by their foundation.

School management structures are not the whole story, though. The R Square of the regressions of school effects on school type dummies is very small, not more than 1%, a finding in line with the school effectiveness literature we cited earlier. There are therefore many other determinants of school quality that, unfortunately, are not in the dataset.

4.5 Long run effects of school quality and peer effects

Specifications 3 and 4 allow for some persistence of the effect of school quality, since past school-grade-year effects are included in the determinants of test scores. In specification 4, we allow for a potential effect of the pupil's background on the progress of the child. The discounting factor therefore measures the long term effect of school quality and of pupil background on progress. These two features matter as long as λ is nonzero.

λ is estimated by minimizing the sum of squared residuals. In specification 3, which include past school-grade-year effects, the decay rate λ is imprecisely estimated. Table 13 shows the sum of squared residuals for a range of λs, from 0 to 0.9. For the 1998-2002 cohort and the 2000-2004 cohort, the optimal discounting factor is zero [10]. For the cohort in-between, the optimal discounting factor is 0.1. But a likelihood ratio test and its associated χ^2 statistic reveal that it is not possible to reject the hypothesis that λ is different from any value between 0 and 0.9[11]. The good news though is that the school-grade-year effects and the pupil effects are robust to small variations of λ around zero[12].

The discounting factor is much more precisely estimated in the last specification (table 14). The

[10] Due to the large number of computations, we decided to estimate λ at a precision greater than 1/10.

[11] Under the null hypothesis that λ is equal to the optimal lambda, i.e. $\lambda^* = 0$ or 0.1, the statistic $2 \cdot ln(L(\lambda)/L(\lambda^* = 0))$ converges to a χ^2 statistic.

[12] Results available on request.

optimal discounting factor is 0 for the 1998-1999 cohort, 0.1 for the 1999-2000 cohort and 0 for the 2000-2004 cohort. Results indicate that on the whole the school-grade-year effect and individual effect specification (equation 2) is not rejected and fits the data as well as the last two specifications. This is evidence that school quality and peer effects may have little long run effects. This is consistent with Hanushek (2003); for instance, reductions of class size have small long term effects (Prais 1996, Krueger 1999). Angrist & Lang (2002) suggest that peer effects in the Boston METCO program were short-lived.

4.6 Pupil Mobility

Mobility is described in table 15. This table does not aim at checking whether the identification assumptions are supported by the data. That analysis will be carried out in the next section. Instead, we wish to provide some stylized facts about pupil mobility in English schools using the estimates of school and pupil effects.

First, pupil mobility in primary school seems on the whole to be a feature of low performing pupils from low quality schools. Pupil effects are indeed negatively correlated with next period school effects and school-grade-year effects. Increasing the pupil effect by 10% of a standard deviation reduces the probability of moving by 0.3%. And increasing the pupil effect by 10% of a standard deviation is correlated with a 0.5 % fall of the next period school-grade-year and a 0.2% fall of the next period school effect. Free school meals are more likely to move (line 5 of table 15). They also move to particular schools, i.e. other schools than most pupils from their school. They move to lower quality school-grade-years (column 3). On the other hand, they move to schools with a higher school fixed effect. This means that they tend to go to schools with a worst peer group but better school quality.

Second, disadvantaged ethnic backgrounds tend to move less than white children and movers from these ethnic backgrounds tend to go to better schools than white children. Bangladeshi pupils are 11.4% less likely to move than white children. Pakistani pupils are 6.8% less likely to move than white children and black carribean children are 5.4% less likely to move. Bangladeshi especially tend to go to better schools, next period school-grade-year effects are higher by 15% of a standard deviation, school effects are higher by 12% of a standard deviation, conditionally on the school effect and school-grade-year effect of grade 2.

5 Robustness Checks and Discussion

In this section, we check whether our identification assumptions are supported by the data. The identification of school quality assumes at least sufficient mobility and exogenous mobility. The identification of peer effects moreover assumes that year-to-year variations in cohort composition are exogenous. We discuss these assumptions in the following subsections.

5.1 Do Children Move Enough to Generate Identification of the Model?

To identify separately the pupil effects from the school or school-grade-year effects, pupils have to move between schools. More precisely the mobility graph as defined in section 2.2 should have one connex component. As we have pointed out, there is only large mobility group covering more than 99.9% of the observations. Table 16 presents some basic statistics on mobility. Most pupils are followed from Key Stage 1 to Key Stage 2. A sizeable proportion (42%) of pupils also change school between grade 2 and grade 6. This is sufficient to generate only one mobility graph. The question of whether students who move are actually different from pupils who do not move is addressed in the following section.

5.2 Is Mobility Endogenous?

School quality and the effect of pupil background on achievement are estimated by comparing pupils' test scores in different schools. It therefore requires that pupil mobility is not driven by unobserved shocks that affect test scores, such as divorce, unemployment, and other family events.

We argue in this section that there is a credibly exogenous source of mobility. Indeed, some primary schools only cater for key stage 1 pupils. Mobility is compulsory in this case. When reestimating the model on compulsory movers only, this paper's results are not significantly affected.

i) Why Endogenous Mobility may be a Problem

We design a small, simple model to understand why endogenous mobility may be a problem. In this model, households experience unemployment shocks that are unobserved by the econometrician. When a household experiences an unemployment shock, children change school and their test scores are likely be lower.

The model is set up as follows. There are two periods. In each period, pupils' parents can either be unemployed $u_{i,t} = 1$, or employed $u_{i,t} = 0$. Test scores are determined by the following equation:

$$y_{i,t} = \theta_i + \psi_{J(i,t)} - \delta u_{i,t} + \eta_{i,t} \qquad (12)$$

(12) is a school effect model. We restrict ourselves to a model with school effects for expositional ease. $y_{i,t}$ is the test score of pupil i in year t. ψ_j is the school effect of school j. δ is the adverse effect of unemployment on test scores, and $u_{i,t}$ is a dummy for unemployment. $\eta_{i,t}$ is a residual.

Unemployment shocks $u_{i,t}$, $i = 1, 2$, $t = 1, 2$, are unobserved and the econometrician estimates the following specification:

$$y_{i,t} = \hat{\theta}_i + \hat{\psi}_{J(i,t)} + \varepsilon_{i,t} \qquad (13)$$

Assuming exogenous mobility, the estimated school effects $\hat{\psi}_j$ are estimated by OLS. To understand the relationship between the structural effects and the least squares estimates, let us write the specification in matrix form.

$$Y = D\theta + F\psi - \delta U + \varepsilon$$

With Y the vector of test scores, D the design matrix for pupil effects, θ the vector of pupil effects, F the design matrix for school effects, ψ the vector of school effects, and ε the residual.

The estimates are as follows:

$$\hat{\theta} = \theta - \delta(D'M_F D)^{-1} D' M_F U \quad (14)$$

$$\hat{\psi} = \psi - \delta(F'M_D F)^{-1} F' M_D U \quad (15)$$

Where M_D is the matrix that projects a vector on the vector space that is orthogonal to D. The same logic applies to M_F.

Thus the estimates of the individual effects and the school effects are biased whenever the correlation between unemployment shocks and design matrices F or D is nonzero, that is, whenever mobility is endogenous. When unobserved unemployment shocks (i) drive pupils to particular schools and (ii) affect their test scores, the estimate of school effects and pupil effects is biased.

ii) Compulsory Moves as an Exogenous Driving Force of Mobility

Back to the real dataset, compulsory movers are children who move between grade 2 and grade 6 because their key stage 1 school does not cater for key stage 2 children. This mobility is likely to be more exogenous than voluntary moves. However, there are three important conditions: (i) compulsory movers should not be significantly different from non compulsory movers; (ii) as compulsory mobility is expected by parents, we need to get evidence that key stage 1 only schools are not particular schools – either better or worse schools; (iii) compulsory mobility provides us with a exogenous reason to move, but it does not per se give an exogenous direction of mobility; children may still sort endogenously into schools.

Table 17 shows descriptive statistics for compulsory movers, noncompulsory movers and stayers. Genders, months of birth, languages and ethnicities are very similar between compulsory movers, noncompulsory movers and stayers. Differences between the three categories appear in the fraction of special needs students and free school meals. The fraction of free school meal students is higher among noncompuslory movers than among compulsory movers, but it very similar between compulsory movers and stayers. On the whole, there are slight differences between the three categories of mobility.

We therefore performed the estimation of specifications 1 and 2 on compulsory movers only[13]. Correlation tables (table 19) reveal that stylized facts are robust to the exclusion of noncompulsory movers: (i) pupil heterogeneity is bigger than school heterogeneity and school-grade-year heterogeneity (ii) the correlation between test scores and individual effects is bigger than the correlation between test scores and either school effects or school-grade-year effects.

Pupil heterogeneity is similar in table 4 and in table 19. School-grade-year or school heterogeneity, while still smaller than pupil heterogeneity, is bigger in the school effects specification with compulsory movers only (6.938 vs 1.941). This might be due to the smaller number of observations in the estimation with compulsory movers only. School effects heterogeneity is comparable in the school-grade-year specification with and without noncompulsory movers. Stylized facts do not change when estimating regressions with compulsory movers only.

The last issue we need to address is whether the direction of mobility is likely to be an identification issue. We define the most frequent school pupils go to. For each school j, the number of pupils who move from school j to school j' is computed. The most frequent school pupils from school j go to in the next

[13]We also performed the estimations of the two other specifications, yielding similar results.

period is noted $M(j)$. Among pupils who move, 63% move to the most frequent school (table 15). This is mainly made up of compulsory movers. Therefore compulsory movers mainly tend to go to the 'usual' school, and the direction of their mobility is not likely to be mainly explained by individual unobserved time varying variables.

5.3 The Identification of Peer Effects: Are Year-to-Year Variations in Grade Composition Exogenous?

The effect of grade composition on school quality is estimated by looking at how year-to-year variations affect school-grade-year effects. This actually requires that year-to-year variations in grade composition are not correlated with other changes in school inputs, such as changes in teacher quality and school funding. One way of addressing this identification issue is to compare year-to-year variations to truly random variations around school average composition.

Formally, if changes in grade composition are truly exogenous, they must be some random fluctuation around the average school composition. In a way identification relies on the idea that grade composition in a given year is a finite size approximation of the school's equilibrium composition.

$$\hat{E}(\mathbf{z}|j,g,t) = E(\mathbf{z}|j,g) + u_{j,g,t} \tag{16}$$

Notations as before. $\hat{E}(\mathbf{z}|j,g,t)$ is the empirical school-grade-year composition in school j, grade g and year t. This is a vector containing the percentage of each ethnicity, the percentage of boys, free meals and special needs. $E(\mathbf{z}|j,g)$ is average school composition across the three cohorts. The size of the noise is approximately normal with variance around $Var\ E(\mathbf{z}|j,g)/\sqrt{n_{j,g,t}}$.

The dataset only contains the empirical composition of grades. Therefore school average composition is just an estimate of the true composition.

$$\hat{E}(\mathbf{z}|j,g) = E(\mathbf{z}|j,g) + v_{j,g,t} \tag{17}$$

With the size of error term approximately $Var\ E(\mathbf{z}|j,g)/\sqrt{n_{j,g}}$. Therefore, finally, $\hat{E}(\mathbf{z}|j,g,t) = \hat{E}(\mathbf{z}|j,g) + u_{j,g,t} - v_{j,g,t}$.

Figure 2 compares the results of simulations to actual year-to-year variations in school-grade-year compositions. For boys, free meals and three important ethnic groups year-to-year variations are remarkably similar to random variations, as in Lavy & Schlosser (2007). This suggest that trends in school-grade-year composition are not likely to explain the results of peer effects regression. On the other hand, year-to-year varations in the fraction of special needs is bigger in the dataset than what would be expected if it were purely random. There may be trends in the fraction of special needs students in schools, which is likely to be due to evolving support for special needs students in English elementary schools. Broadly speaking, apart from special needs students, variations in gender and ethnic compositions are similar to random variations around average school composition.

5.4 What do League Tables Tell Us?

The Education Reform Act 1988 set up the National Curriculum, which follows pupils through key stages, as we pointed out in section 3.1. Measures of performance at the end of each key stage are now disclosed on the BBC's website and through local newspapers. This is a crucial element of transparency that is coupled with some leeway for school choice. Parents typically submit their first three choices to Local Education Authorities in the fall of the academic year before enrollment; most faith schools require a special application form.

Measures of performance are published in league tables. Typing the postcode of a school on the website of the BBC will reveal three pieces of information: (i) the average test score at key stage examinations; (ii) the average value added of pupils in the school; value added is the difference between the pupil's test score in the previous key stage and his current test score; finally, (iii) the average test score and value added in the local authority.

Are these elements informative about school effectiveness? The answer depends on the shape of the production function. It turns out that, using our models, neither absolute test scores nor value added measures are good estimates of school effectiveness φ or ψ. In the full-fledged model with non-zero decay rate and past inputs, the average test score of a school is a mixture of school effectiveness, the average individual effect in the school and the average effectiveness of past schools. Indeed,

$$E[y_{i,f,t}|j,g=6,t] = (1+\lambda)\cdot E[\theta_i|j,g=6,t] + \varphi_{j,g=6,t} + \lambda \cdot E[\varphi_{j,g=2,t-4}|j,g=2,t-4] \tag{18}$$

Where notations are as before. t is the year, $g=6$ says that we are considering test scores in grade 6. $E[y_{i,f,2}|j,g=6,t]$ is the average test score in school j, grade g in grade 6. In two of the three cohorts, the estimated decay rate is not different from zero. In this case, average past school effectiveness disappears but the average individual effect remains. Therefore, unless two schools have the same intake, the average test score is not informative about φ.

How important is the contribution of individual fixed effects to school average test scores? Table 20 shows the decomposition of the between-school variance of test scores into its components. Most of the variance of pupil effects is within schools (76%). There is however substantial within-school variance of the pupil effects (24% of the variance of pupil effects). More troubling, the between-school variance of individual effects is very close to the between-school variance of test scores. This suggests that average test scores are a flawed measure of school effectiveness, provided our specification is correct.

Value added measures are a means to get rid of these confounding effects. Average value added is:

$$E[\Delta y_{i,f}|j,g=6,t] = \lambda \cdot E[\theta_i|j,g=6,t] + \varphi_{j,g=6,t} + (\lambda-1)\cdot E[\varphi_{j,g=2,t-4}|j,g=2,t-4] \tag{19}$$

Where $\Delta y_{i,f} = y_{i,f,2} - y_{i,f,1}$. $E[\Delta y_{i,f}|j,t]$ is average value added in school j, in a given year t. Again, λ is close to zero in most cohorts, so that average value added is free of the individual effects. However, average past school quality still enters the equation. It is a priori a problem since the variance of school effects in grade 2 is comparable to the variance of school effects in grade 6.

Overall, neither average test scores nor average value added measures are a good proxy for school effectiveness. It seems that more elaborate statistics are needed to truly help parents in their school choice.

5.5 The Matching of Pupils to Schools

Results have shown that the most relevant specification is equation 2. In this specification, schools are equally effective for all students, that is, there is no complementarity between pupils and schools. If the educational production function is truly specified as in specification 2, the model predicts that there will be no matching between pupil effects and school effects at equilibrium, following Becker (1973). Matching patterns are indeed determined by the complementarity between pupil effects and school effects. [14] In such a world, the model predicts zero correlation between pupil effects and school-grade-year fixed effects.

However, some of the correlations between child effects and school effects in table 4 and in table 5 are negative. Does it mean that pupils with a high pupil effect are structurally matched with low school-grade-year effects? The correlation between estimated pupil effects and estimated school effects is actually downward biased and we perform simulations to estimate the magnitude of the bias, suggesting that the correlation is likely to be close to zero.

The correlation between estimated effects is downward biased. This has been pointed out in the context of worker-firm matched panel datasets (Abowd & Kramarz 2004). To make this clear, let us decompose the correlation between child and school effects. This correlation can be written as the sum of the correlation between measurement errors and the true covariance between the effects.

$$Cov(\hat{\theta}, \hat{\varphi}) = Cov(\hat{\theta} - \theta, \hat{\varphi} - \varphi) + Cov(\hat{\theta} - \theta, \varphi) + Cov(\hat{\varphi} - \varphi, \theta) + Cov(\theta, \varphi)$$

θ is the individual effect, φ is the school-grade-year effect, $\hat{\theta}$ is the estimated individual effect, $\hat{\varphi}$ is the estimated school-grade-year effect.

The estimation of $Cov(\theta, \varphi)$ therefore requires the estimation of $Cov(\hat{\theta} - \theta, \varphi)$, $Cov(\varphi, \hat{\theta} - \theta)$, $Cov(\hat{\theta} - \theta, \hat{\varphi} - \varphi)$. In general, the measurement errors of child and school effects are negatively correlated (Abowd & Kramarz 2004). The intuition behind this result is that (i) pupils who change school get a better estimated effect but school effects are less precisely estimated (ii) pupils who do not change school have a less well estimated effect but their associated school effect is more precisely estimated.

Simulations can assess the order of magnitude of the downward bias of the correlation. We generate pupil effects who have a normal distribution with the same variance as the estimated pupil effects. We also generate school effects the same way. The point here is that pupil effects and school effects are uncorrelated. We then generate simulated test scores using the specification with past and current school-grade-year and individual effects.

Results are presented in table 21. Simulated pupil effects and school-grade-year effects were generated using the variances of the last table of table 5. The results of simulations suggest that even in the absence of a true correlation between pupil effects and school effects, the correlation between estimated effects is negative. The correlation is -0.033. The empirical correlation is stable across the three simulations. The correlation between school-grade-year and individual effects is therefore likely to be close to -0.1, with school-grade-year effects explaining little of the variance of test scores.

Generally speaking, estimating the correlation between pupil effects and school effects is a difficult challenge in pupil-school or worker-firm fixed effects specifications. Most papers find a zero or negative

[14] This of course, assumes a particular form of preferences and special market conditions. The housing market should be perfect, parents should know the educational production function as specified in equation 2 and the only reason for location decisions should be the level of test scores.

correlation (Abowd & Kramarz 2004, Abowd et al. 1999). But these papers do not include a match effect that could account for the complementarity between pupil and school-grade-year effects or worker and firm effect. The identification proofs of appendix A are not valid in this case and more stringent identification assumptions are needed (Woodcock 2007).

6 Conclusion

This aim of this paper has been to use rich data on English primary school children, and the schools they attend, to estimate the respective contributions of pupil background, peers and schools in educational achievement. Two ingredients were needed. First, a comprehensive English dataset provides us with the test scores of all English pupils in state schools at age 7 and 11, linking pupils to their school and their peers. And second, we develop a set of econometric techniques that allow us to decompose test scores into the effect of the background, the effect of the peers and the effect of the schools. This is identified under general conditions of sufficient and exogenous mobility.

The main finding is that, from this detailed econometric model, pupil ability and background is probably the most important educational input in that it explains a large fraction of the overall variance. This suggests that either inherited ability, early educational experiences acquired before the age of seven) and family background play a very important part in the educational process. School time-invariant inputs are the second most important input but far less important than pupil effects, provided the identification and specification of our models is correct. Peer effects may be the least most important input, most effects being small.

The analysis of mobility reveals that a substantial fraction of mobility is due to the structure of the English schools system where compulsory mobility occurs at certain stages. This provides us with a reasonably exogenous source of mobility. Results reveal that high achieving pupils either tend to stay in the same school, or to go to the most usual school which students go to.

The findings of this paper should be useful to a number of audiences and research areas. First of all, it has clear implications for education policy and design. It is evident that pupil-specific factors matter most, a conclusion also reached in other areas of the social sciences, through very different modelling approaches (e.g. the 'school effectiveness' literature in education research). Second, this paper applies and refines methods from the literature on matched worker-firm data to pupils in schools. We think this is important, providing strong evidence showing how economic agents (in this case pupils) behave and adapt in the environment (in this case schools) in which they spend a considerable amount of their time. It is very clear from our analysis (and from related ones like Rivkin et al. (2005)) that one gains a lot from observing pupils in the schools they attend and we can say a lot more than a study of educational achievement based only on pupils or only on schools is able to. Thirdly, and more generally speaking, this paper estimates a full-blown specification with semi-parametric estimates of pupil background effects, peer effects and school effects. It is likely to be the first paper of this kind to include peer-effects and to include the effect of past schools(-grade-year) as current inputs and the findings broadly confirm the initial claim of the Coleman report. These conclusions may be of substantial interest to policy makers seeking to spend public funding into the right inputs either to increase efficiency or to narrow educational inequalities.

References

Abowd, J., Creecy, R. & Kramarz, F. (2002), Computing person and firm effects using linked longitudinal employer-employee dataset.

Abowd, J. M. & Kramarz, F. (2004), Are good workers employed by good firms? a simple test of positive assortative matching models, Econometric Society 2004 North American Winter Meetings 385, Econometric Society.

Abowd, J. M., Kramarz, F. & Margolis, D. N. (1999), 'High wage workers and high wage firms', *Econometrica* **67**(2), 251–334.

Angrist, J. D. & Lang, K. (2002), How important are classroom peer effects? evidence from boston's metco program, Staff Working Papers 02-85, W.E. Upjohn Institute for Employment Research.

Becker, G. S. (1973), 'A theory of marriage: Part i', *Journal of Political Economy* **81**(4), 813–46.

Blundell, R. & Robin, J. M. (1999), 'Estimation in large and disaggregated demand systems: An estimator for conditionally linear systems', *Journal of Applied Econometrics* **14**(3), 209–32.

Bryk, A. & Raudenbush, S. (1992), *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage, Newbury Park.

Coleman, J. (1966), *Equality of Educational Opportunity*, U.S. GPO, Washington, D.C.

Currie, J. (2001), 'Early childhood education programs', *Journal of Economic Perspectives* **15**(2), 213–238.

Dearden, L., Ferri, J. & Meghir, C. (2002), 'The effect of school quality on educational attainment and wages', *The Review of Economics and Statistics* **84**(1), 1–20.

Dongarra, J., Duff, I., Sorensen, D. & van der Vorst, H. (1991), *Solving Linear Systems on Vector and Shared Memory Computers*, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, United States.

Garces, E., Thomas, D. & Currie, J. (2002), 'Longer-term effects of head start', *American Economic Review* **92**(4), 999–1012.

Gibbons, S., Machin, S. & Silva, O. (forthcoming), 'Choice, competition and pupil achievement', *Journal of the European Economic Association* .

Goldstein, H. (2002), *Multilevel Statistical Models*, Hodder Arnold.

Gould, E. D., Lavy, V. & Paserman, M. D. (2004), Does immigration affect the long-term educational outcomes of natives? quasi-experimental evidence, NBER Working Papers 10844, National Bureau of Economic Research, Inc.

Hanushek, E. A. (1986), 'The economics of schooling: Production and efficiency in public schools', *Journal of Economic Literature* **XXIV**, 1141–1177.

Hanushek, E. A. (2003), 'The failure of input-based schooling policies', *Economic Journal* **113**(485), F64–F98.

Hanushek, E. A. & Rivkin, S. G. (2003), Does public school competition affect teacher quality?, *in* C. M. Hoxby, ed., 'The Economics of School Choice', National Bureau of Economic Research, chapter 1.

Heckman, J. & Masterov, D. V. (2007), 'The productivity argument for investing in young children', *Review of Agricultural Economics* **29**(3), 446–493.

Hoxby, C. (2000), Peer effects in the classroom: Learning from gender and race variation, NBER Working Papers 7867, National Bureau of Economic Research, Inc.

Krueger, A. B. (1999), 'Experimental estimates of education production functions', *The Quarterly Journal of Economics* **114**(2), 497–532.

Lavy, V. & Schlosser, A. (2007), Mechanisms and impacts of gender peer effects at school, NBER Working Papers 13292, National Bureau of Economic Research, Inc.

Levacic, R. & Vignoles, A. (2002), 'Researching the links between school resources and student outcomes in the uk: A review of issues and evidence', *Education Economics* **10**(3), 313–331.

Machin, S. & Vignoles, A. (2005), *What's the Good of Education? The Economics of Education in the United Kingdom*, Princeton University Press, Princeton, New Jersey, United States.

Manski, C. F. (1993), 'Identification of endogenous social effects: The reflection problem', *Review of Economic Studies* **60**(3), 531–42.

Manski, C. F. (2000), 'Economic analysis of social interactions', *Journal of Economic Perspectives* **14**(3), 115–136.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D. & Ecob, R. (1988), *School Matters: The Junior Years*, Paul Chapman.

Prais, S. (1996), 'Class-size and learning: the tennessee experiment – what follows?', *Oxford Review of Education* .

Rao, C. & Sinharay, S. (2006), *Handbook of Statistics: Psychometrics*, Elsevier, Amsterdam.

Raudenbush, S. & Bryk, A. S. (1986), 'A hierarchical model for studying school effects', *Sociology of Education* **59**(1), 1–17.

Rivkin, S. G., Hanushek, E. A. & Kain, J. F. (2005), 'Teachers, schools, and academic achievement', *Econometrica* **73**(2), 417–458.

Stiefel, L., Schwartz, A., Rubenstein, R. & Zabel, J. (2005), *Measuring School Performance and Efficiency: Implications for Practice and Research*, Eye on Education.

Summers, A. A. & Wolfe, B. L. (1977), 'Do schools make a difference?', *American Economic Review* **67**(4), 639–52.

Teddlie, C. & Reynolds, D. (1999), *The International Handbook of School Effectiveness Research*, Routledge Falmer.

Todd, P. E. & Wolpin, K. I. (2003), 'On the specification and estimation of the production function for cognitive achievement', *Economic Journal* **113**(485), F3–F33.

West, A. & Pennell (2003), *Underachievement in schools*, Routledge Falmer.

Woodcock, S. (2007), Match effects. unpublished manuscript.

Wooldridge, J. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, Mass.

A Appendix: Identification of The Current and Past School Effects Model

This section gives sufficient conditions for the identification of the student effects, school effects and past school effects. The identification condition is an extension of Abowd et al. (2002).

We will start with the identification of the model and we will consider a continuous distribution of students with a finite number of schools. A random mobility graph describes the movement of individuals between units. Units are edges, individual mobility is represented by vertices. A transition probability between two schools is defined as the probability that an individual belongs to the two units in two subsequent time periods.

A.1 Notations

There are T time periods $t = 1, \dots, T$. Individuals are indexed by $i = 1, \dots, N$, and units are indexed by $j \in \mathcal{J} = \{1, \dots, J\}$. I write that $i \in j$ if individual i has belonged to unit j in one of the periods $t = 1, \dots, T$.

We build a graph connecting schools in the following way. There is a connection between school j and school j' if the probability that a pupil belonged to both school j and school j' is strictly positive. Formally, $(\mathcal{J}, G)$ is called a mobility graph with $\mathcal{J}$ the set of schools and G the set of vertices defined as follows:

$$\{j, j'\} \in G \quad \Leftrightarrow \quad P(\{i \in [0, N], \quad i \in j, i \in j'\}) > 0$$

A.2 Identification of the fixed effects conditionally on λ

I prove the following sufficient condition for the identification of school fixed effects. Let $\jmath \in \mathcal{J}$ be an arbitrary school and $\imath$ an individual who belonged to school $\jmath$. By convention, we set $\psi_\jmath = 0$ and $\theta_\imath = 0$.

Theorem 1 *The school effects ψ_k of the connex component of $(\mathcal{J}, G)$ containing unit $\jmath$ are jointly identifiable. The pupil effects of the individuals of the vertices of this connex component are all jointly identifiable.*

Proof It is sufficient to show that, for any pair of schools j and j', if ψ_j is identified, and j and j' are connected, then $\psi_{j'}$ is identified.

If j and j' are connected, one of the following equalities is satisfied:

$$E_i[(y_{i,t+1} - \lambda y_{i,t}) - (y_{i,t} - \lambda y_{i,t-1}) | J(i, t+1) = j, J(i, t) = j'] = \lambda(\psi_j - \psi_{j'}) \tag{A-20}$$

or

$$E_i[(y_{i,t+1} - \lambda y_{i,t}) - (y_{i,t} - \lambda y_{i,t-1}) | J(i, t+1) = j', J(i, t) = j] = \lambda(\psi_{j'} - \psi_j) \tag{A-21}$$

Setting $y_{i,0} = 0$ by convention. It is clear from this relationship that when $\lambda \neq 0$, $\psi_{j'}$ is identified implies that ψ_j is identified.

Since ψ_j is identified, and, for all j, j' whenever ψ_j is identified and j and j' are connected, $\psi_{j'}$ is identified, then by recursion the connex component containing ψ_j is identified.

I now turn to the identification of individual fixed effects. When a student connects two schools whose school effects are identifiable, then the corresponding student effect is identifiable. This shows that all student effects of the vertices of the connex component of ψ_j are identified. □

A.3 The identification of λ

Previous sections have given the conditons for the identification of the fixed effects when λ is given. In this section, we assume that the mobility graph $(\mathcal{J}, G)$ of the dataset has one connex component, and we put forward an iterated estimator for λ. Let $\hat{\beta}(\lambda)$, $\hat{\theta}(\lambda)$, $\hat{\psi}(\lambda)$ be the estimates of the fixed effects conditionally on λ. y can be written as:

$$y_{i,t} = \mathbf{x_{i,t}}\hat{\beta}(\lambda) + \sum_{k=0}^{t} \lambda^k \hat{\theta}_i(\lambda) + \sum_{k=0}^{T} \lambda^k \hat{\psi}_{J(i,t-k)}(\lambda) + \hat{\varepsilon_{i,t}}(\lambda)$$

A potential estimator for λ is given by maximizing the likelihood of the model conditionally on the estimated pupil effects $\hat{\theta}_i(\lambda)$ and the school effects $\hat{\psi}_i(\lambda)$. This assumes that errors are normally distributed and orthogonal across observations. The proof of identification for λ can be found in Blundell & Robin (1999).

B Appendix: Estimation of the Model with Current and Past School and Pupil Effects (Specification 4)

B.1 Matrix formulation of the Model

I write the specification in matrix form to get the normal form equations and proceed to the estimation by conjugate gradient. The number of different students is N. The number of schools is J. The number of covariates is K. The number of observations in the ith period is n_i, and n is the total number of observations, $n = \sum_{t=1}^{T} n_t$.

$$\begin{array}{rcl} Y & = & X\beta + D_\lambda\theta + \Phi_\lambda\psi + U \\ \text{with} & & \Phi_\lambda = F + \lambda F_{-1} + \cdots + \lambda^T F_{-T} \end{array} \tag{B-22}$$

Observations are ordered such that the vector of observations Y contains observations of the first period, then the observations of the second period.

$$Y = (Y_1'|Y_2'|\cdots|Y_T')' \qquad Y \in \mathbb{R}^n$$

Y_i is a column vector with n_i elements. The design matrix of individuals D is a matrix of $n \times N$ elements. Again, I decompose the design matrix into a first period matrix and a second period matrix.

$$D = (D_1'|(1+\lambda)D_2'|\cdots|\sum_{k=0}^{T}\lambda^k D_k')' \qquad D \in \mathcal{M}_{n,N}(\mathbb{R})$$

The design matrices for units are linked.

$$F = (G_1'|G_2'|\cdots|G_T')' \qquad F \in \mathcal{M}_{n,J}(\mathbb{R})$$

$$F_{-1} = (0|G_1^{-1'}|G_2^{-1'}|\cdots|G_{T-1}^{-1'})'$$
$$F_{-2} = (0|0|G_1^{-2'}|G_2^{-2'}|\cdots|G_{T-2}^{-2'})'$$

With these notations the identification hypothesis – namely that no unobserved time-varying shock should be correlated with the covariates, student and/or school effects – can be translated in matrix form, ie:

$$E(U|D_\lambda, \Phi_\lambda, X) = 0 \tag{B-23}$$

The normal equations follow:

$$\begin{aligned} X'(Y - X\beta - D_\lambda\theta - \Phi_\lambda\psi) &= 0 \qquad \text{(B-24)} \\ D'(Y - X\beta - D_\lambda\theta - \Phi_\lambda\psi) &= 0 \qquad \text{(B-25)} \\ \Phi_\lambda'(Y - X\beta - D_\lambda\theta - \Phi_\lambda\psi) &= 0 \qquad \text{(B-26)} \end{aligned}$$

Which can be written:

$$A_\lambda b = \begin{pmatrix} X'X & X'D_\lambda & X'\Phi_\lambda \\ D_\lambda'X & D_\lambda'D_\lambda & D_\lambda'\Phi_\lambda \\ \Phi_\lambda'X & \Phi_\lambda'D_\lambda & \Phi_\lambda'\Phi_\lambda \end{pmatrix} \begin{pmatrix} \beta \\ \theta \\ \psi \end{pmatrix} = \begin{pmatrix} X'Y \\ D_\lambda'Y \\ \Phi_\lambda'Y \end{pmatrix} \tag{B-27}$$

b is the parameter vector. A_λ is non-singular under conditions exposed in section 2. Moreover A_λ is a symmetric, positive definite matrix, and therefore the problem of finding b can be solved by the conjugate gradient algorithm.

Let us denote by $n_{i,j,t}$ the number of observations of student i in school j in year t.

$$\begin{aligned} D_\lambda'D_\lambda &= Diag(n_{1,\cdot,\cdot} + \lambda n_{1,\cdot,1}, n_{2,\cdot,\cdot} + \lambda n_{2,\cdot,1}, \ldots, n_{N,\cdot,\cdot} + \lambda n_{N,\cdot,1}) \\ F'F &= Diag(n_{\cdot,1,\cdot}, n_{\cdot,2,\cdot}, \ldots, n_{\cdot,J,\cdot}) \\ F_{-1}'F_{-1} &= Diag(n_{\cdot,1,1}, n_{\cdot,2,1}, \ldots, n_{\cdot,J,1}) \end{aligned}$$

And,

$$\Phi_\lambda'\Phi_\lambda = F'F + \lambda F'F_{-1} + \lambda F_{-1}'F + \lambda^2 F_{-1}'F_{-1}$$

$$
\begin{aligned}
D'F &= [n_{i,j,\cdot}]_{i,j} & D'F_{-1} &= [n_{i,j,1}]_{i=1,\dots,N,j=1,\dots,J} \\
F'F_{-1} &= (0|G_2'G_1) & G_2'G_1 &= [m_{j,j'}]_{j,j'=1,\dots,J}
\end{aligned}
$$

Where $m_{j,j'}$ is the number of observations that move from unit j to unit j' between period 1 and 2.

The estimation of the normal form equations uses a conjugate gradient estimator. Recursive sequence $b_n' = (\beta_n' \quad \theta_n' \quad \psi_n')$ is defined such that $b_0 = 0$, and b_n is built from b_{n-1} by the conjugate gradient algorithm described in Dongarra et al. (1991).[15]

B.2 Estimating the discounting factor λ

The estimation of λ proceeds in the following way. Conditionally on the discounting factor λ, previous sections have shown that I get estimates of the effects $(\hat{\theta}_i(\lambda), i = 1 \dots N, \hat{\psi}_j(\lambda), j = 1 \dots J, \hat{\beta}(\lambda))$ by OLS. Two methods are feasible. Firstly, the model is linear conditionally on λ, and therefore the identification method described in Blundell & Robin (1999) applies. Secondly, it is possible estimate the parameter by minimizing the residual sum of squares – or maximizing likelihood estimation.

[15]The conjugate gradient is mathematically an exact method. Due to rounding errors in the computational process, it is practically an approximation of the true solution. The speed of convergence depends on the condition number of matrix A_λ, ie the ratio of its highest eigenvalue and its lowest eigenvalue. Reducing the condition number increases the convergence speed. This is the purpose of preconditioning, described in the appendix.

Number of pupils	1,783,281			
Number of schools	20,708			
Number of School-Grade-Years	96,163			

	Number	Percentage
Sample size	8,660,561	(100.00 %)
Key Stage 1 Observations	3,403,213	(39.30%)
. . . 1998	1,122,020	(12.96%)
. . . 1999	1,158,427	(13.38%)
. . . 2000	1,122,751	(12.96%)
Key Stage 2 Observations	5,257,348	(60.70%)
. . . 2002	1,703,250	(19.67%)
. . . 2003	1,821,894	(21.04%)
. . . 2004	1,732,219	(20.00%)

	Number	Percentage
Male	4,413,066	(0.51)
Free School Meal	1,486,517	(0.17)
Special Needs	1,966,563	(0.23)
English spoken at home	7,893,062	(0.91)

	Mean	(Std. Dev.)	Min.	Max.
All Test Scores	49.89	(9.96)	12.68	80.57
Key Stage 1 Test Scores	49.95	(9.95)	12.93	80.57
English	49.94	(9.95)	21.02	80.57
Maths	49.96	(9.95)	12.93	78.00
Key Stage 2 Test Scores	49.86	(9.96)	12.68	74.26
English	49.82	(9.96)	20.48	74.13
Maths	49.86	(9.98)	23.99	66.45
Science	49.88	(9.95)	12.68	74.26

Table 2: Descriptive Statistics

Type	Faith	Governors	Admissions Authority	Assets Owned By	Employer
Non-Majority Controlled Schools					
Community Schools	Secular	Parents > 30 % LEA 20% Staff < 30% Community 20%	LEA	LEA	LEA
Voluntary Controlled	Mostly Church of England, Some other faiths Some secular	Foundation < 25% Parents >30% LEA < 20% Staff < 30% Community 10%	LEA	LEA	LEA
Foundation	Mostly secular some Church of England	Foundation < 25% Parents > 30% LEA < 20% Staff <30% Community 10%	Governors	Curch or charity	Governors
Majority Controlled Schools					
Voluntary Aided	Mostly Church of England Catholic, some secular	Foundation > 50 % Parents > 30% LEA <10% Staff <30%	Governors	Church or Charity	Governors

This table is taken from Gibbons, Machin & Silva (forthcoming)

Table 3: Primary School Categories in England

– School and Pupil Fixed Effects –

	Mean	Std. Dev.	y	θ	θ^*	$\theta^\perp$	ψ
y, Standardized grade	49.893	9.958	1.000				
θ, Pupil Effect	-0.000	9.300	0.814	1.000			
θ^*, Observed Pupil Heterogeneity	-0.000	6.333	0.481	0.681	1.000		
$\theta^\perp$, Unobserved Pupil Heterogeneity	0.000	6.810	0.664	0.732	0.000	1.000	
ψ, School Effect	0.045	1.941	0.121	-0.076	-0.001	-0.103	1.000
ε, Residual	0.000	4.634	0.465	0.000	-0.000	0.000	0.000

– School Grade Year Effect and Pupil Effect –

	Mean	Std. Dev.	y	θ	θ^*	$\theta^\perp$	φ	ψ
y, Standardized grade	49.893	9.958	1.000					
θ, Pupil Effect	-0.000	9.473	0.791	1.000				
θ^*, Observed Pupil Heterogeneity	-0.000	6.471	0.471	0.683	1.000			
$\theta^\perp$, Unobserved Pupil Heterogeneity	-0.000	6.918	0.642	0.730	0.000	1.000		
φ, School-Grade-Year Effect	13.811	7.446	0.074	0.153	0.298	-0.070	1.000	
ψ, School Effect	-0.000	2.176	0.157	-0.062	0.002	-0.087	0.293	1.000
ε, Residual	0.000	4.470	0.449	0.000	-0.000	0.000	0.000	0.000

Estimated with a2group, a2reg, xtlreg and xtlreg2. Available through the corresponding author Amine Ouazad.
Reading: Test Scores have a standard deviation of 10 and a mean of 50.
Source: National Pupils Database, Department for Education and Skills.

Table 4: Correlation Tables

– Past and Current School Grade Year Effect, Current Individual Effect –

	Mean	Std. Dev.	y	θ	θ^*	$\theta^\perp$	φ	ψ
y, Standardized grade	49.893	9.958	1.000					
θ, Pupil Effect	0.000	8.922	0.833	1.000				
θ^*, Observed Pupil Heterogeneity	0.000	5.442	0.547	0.610	1.000			
$\theta^\perp$, Unobserved Pupil Heterogeneity	0.000	7.071	0.631	0.793	-0.000	1.000		
φ, School-Grade-Year Effect	1.427	3.932	0.117	-0.153	0.004	-0.196	1.000	
ψ, School Effect	0.000	2.309	0.131	-0.114	-0.015	-0.133	0.584	1.000
ε, Residual	-0.000	4.435	0.445	0.000	-0.000	0.000	-0.000	-0.000

– Past and Current School Grade Year and Individual Effects –

	Mean	Std. Dev.	y	θ	θ^*	$\theta^\perp$	φ	ψ
y, Standardized grade	49.893	9.958	1.000					
θ, Pupil Effect	-0.000	8.737	0.833	1.000				
θ^*, Observed Pupil Heterogeneity	-0.000	5.331	0.547	0.610	1.000			
$\theta^\perp$, Unobserved Pupil Heterogeneity	-0.000	6.922	0.630	0.792	-0.000	1.000		
φ, School-Grade-Year Effect	1.154	4.445	0.107	-0.138	0.003	-0.177	1.000	
ψ, School Effect	-0.000	2.306	0.125	-0.120	-0.017	-0.138	0.516	1.000
ε, Residual	0.000	4.437	0.451	0.000	0.001	-0.000	0.000	-0.000

Estimated with a2group, a2reg, xtlreg and xtlreg2. Available through the corresponding author Amine Ouazad.
Reading: Test Scores have a standard deviation of 10 and a mean of 50.
Source: National Pupils Database, Department for Education and Skills.

Table 5: Correlation Tables

– School and Pupil Fixed Effects –

	Mean	Std. Dev.	y	θ	θ^*	$\theta^\perp$	ψ
y, Standardized grade	49.893	9.958	1.000				
θ, Pupil Effect	-0.000	9.300	75.337	1.000			
θ^*, Observed Pupil Heterogeneity	-0.000	6.333	30.321	40.111	1.000		
$\theta^\perp$, Unobserved Pupil Heterogeneity	0.000	6.810	45.016	46.374	0.000	1.000	
ψ, School Effect	0.045	1.941	2.340	-1.375	-0.018	-1.357	1.000
ε, Residual	0.000	4.634	21.474	0.000	-0.000	0.000	0.000

– School Grade Year Effect and Pupil Effect –

	Mean	Std. Dev.	y	θ	θ^*	$\theta^\perp$	φ	ψ
y, Standardized grade	49.893	9.958	1.000					
θ, Pupil Effect	-0.000	9.473	74.600	1.000				
θ^*, Observed Pupil Heterogeneity	-0.000	6.471	30.347	41.873	1.000			
$\theta^\perp$, Unobserved Pupil Heterogeneity	-0.000	6.918	44.253	47.861	0.000	1.000		
φ, School-Grade-Year Effect	13.811	7.446	5.474	10.761	14.369	-3.607	1.000	
ψ, School Effect	0.000	2.723	7.680	0.431	1.402	-0.971	7.440	1.000
ε, Residual	0.000	4.470	19.980	0.000	-0.000	0.000	0.000	-0.015

Estimated with a2group, a2reg, xtlreg and xtlreg2. Available through the corresponding author Amine Ouazad.
Reading: Test Scores have a standard deviation of 10 and a mean of 50.
Source: National Pupils Database, Department for Education and Skills.

Table 6: Covariance Tables

– Past and Current School Grade Year Effect, Current Individual Effect –

	Mean	Std. Dev.	y	θ	θ^*	$\theta^\perp$	φ	ψ
y, Standardized grade	49.893	9.958	1.000					
θ, Pupil Effect	0.000	8.922	74.037	1.000				
θ^*, Observed Pupil Heterogeneity	0.000	5.442	29.615	29.610	1.000			
$\theta^\perp$, Unobserved Pupil Heterogeneity	0.000	7.071	44.421	50.001	-0.000	1.000		
φ, School-Grade-Year Effect	1.427	3.932	4.569	-5.369	0.080	-5.448	1.000	
ψ, School Effect	-0.000	3.020	4.216	-4.917	0.146	-5.063	9.031	1.000
ε, Residual	-0.000	4.435	19.668	0.000	-0.000	0.000	-0.000	0.034

– Past and Current School Grade Year and Individual Effects –

	Mean	Std. Dev.	y	θ	θ^*	$\theta^\perp$	φ	ψ
y, Standardized grade	49.893	9.958	1.000					
θ, Pupil Effect	-0.000	8.737	72.474	1.000				
θ^*, Observed Pupil Heterogeneity	-0.000	5.331	29.017	28.424	1.000			
$\theta^\perp$, Unobserved Pupil Heterogeneity	-0.000	6.922	43.457	47.913	-0.000	1.000		
φ, School-Grade-Year Effect	1.154	4.445	4.741	-5.368	0.071	-5.439	1.000	
ψ, School Effect	-0.000	3.014	3.961	-5.031	0.119	-5.150	8.993	1.000
ε, Residual	0.000	4.437	19.682	0.004	0.018	-0.014	0.000	0.010

Estimated with a2group, a2reg, xtlreg and xtlreg2. Available through the corresponding author Amine Ouazad.
Reading: Test Scores have a standard deviation of 10 and a mean of 50.
Source: National Pupils Database, Department for Education and Skills.

Table 7: Covariance Tables

	Dependent variable: Pupil Fixed Effect Sample: Key Stage 1 and Key Stage 2 Specification			
	School Effects	School-Grade-Year f.e.	Past S.-G.-Y. f.e.	Past Individual f.e.
Male	0.259**	0.256**	0.254**	0.255**
	(0.010)	(0.010)	(0.011)	(0.010)
Free School Meal	-4.097**	-4.080**	-4.067**	-3.987**
	(0.015)	(0.015)	(0.015)	(0.015)
Special Needs	-11.268**	-11.238**	-11.287**	-11.064**
	(0.014)	(0.014)	(0.014)	(0.014)
Month Of Birth	-0.274**	-0.276**	-0.276**	-0.268**
	(0.001)	(0.001)	(0.001)	(0.001)
Chinese	1.633**	1.567**	1.644**	1.622**
	(0.094)	(0.096)	(0.097)	(0.095)
Mixed	0.449**	0.395**	0.426**	0.415**
	(0.039)	(0.040)	(0.041)	(0.039)
Indian	-0.643**	-0.777**	-0.755**	-0.733**
	(0.034)	(0.035)	(0.035)	(0.034)
White	Ref.	Ref.	Ref.	Ref.
Bangladeshi	-3.124**	-3.456**	-3.429**	-3.350**
	(0.058)	(0.059)	(0.059)	(0.058)
Black African	-1.794**	-2.004**	-1.930**	-1.882**
	(0.048)	(0.048)	(0.049)	(0.047)
Pakistani	-4.028**	-3.993**	-3.970**	-3.874**
	(0.036)	(0.036)	(0.036)	(0.035)
Black, Other	-0.855**	-1.018**	-0.965**	-0.957**
	(0.073)	(0.074)	(0.075)	(0.073)
Other ethnicity	-0.525**	-0.587**	-0.556**	-0.529**
	(0.027)	(0.027)	(0.028)	(0.027)
Black Carribean	-1.540**	-1.631**	-1.571**	-1.548**
	(0.045)	(0.045)	(0.045)	(0.044)
R Squared	0.46	0.47	0.37	0.37
F Statistic	72,860.81	74,926.46	49,801.00	49,851.24
Number of Pupils	1,783,255	1,783,255	1,783,255	1,783,255

Source: National Pupils Database, Department for Education and Skills.
**: Significant at 1%. *: Significant at 5%.
Reading: Test Scores have a standard deviation of 10 and a mean of 50.

Table 8: Analysis of Pupil Fixed Effects

	Dependent variable: $\varphi_{j,g,t}$ School-Grade-Year Effect		
Specification	School-Grade-Year f.e.	Past S.-G.-Y. f.e.	Past Individual f.e.
Fraction in Grade			
Male	-0.445**	-0.395*	-0.390*
	(0.157)	(0.181)	(0.178)
Free School Meal	0.678**	0.536*	0.458
	(0.236)	(0.257)	(0.252)
Special Needs	-0.171	-0.329	-0.586**
	(0.175)	(0.198)	(0.195)
White	Ref.	Ref.	Ref.
Chinese	-2.063	-3.412	-3.233
	(1.685)	(1.790)	(1.766)
Mixed	-0.873	-0.954	-1.066
	(0.541)	(0.586)	(0.577)
Indian	-0.490	-0.715	-0.619
	(0.801)	(0.857)	(0.849)
Bangladeshi	-1.802	-1.624	-1.481
	(1.763)	(1.781)	(1.773)
Black African	-1.205	-0.830	-0.733
	(0.886)	(0.934)	(0.927)
Pakistani	0.521	0.350	0.470
	(0.818)	(0.860)	(0.844)
Black, Other	1.903	1.637	1.750
	(1.020)	(1.085)	(1.070)
Black Carribean	0.649	0.467	0.593
	(0.841)	(0.882)	(0.875)
School Fixed Effects	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes
R Squared	0.87	0.75	0.79
F Statistic	10,814.15	3,956.89	5,686.27
Number of observations	8,660,468	8,660,468	8,660,468
Number of school-grade-years	96,154	96,154	96,154
Number of schools	20,705	20,705	20,705

Source: National Pupils Database, Department for Education and Skills.
**: Significant at 1%. *: Significant at 5%.
Reading: Test Scores have a standard deviation of 10 and a mean of 50.

Table 9: Peer Effects in Schools in Grade 2 – Analysis of School-Grade-Year Effects

	Dependent variable: $\varphi_{j,g,t}$ School-Grade-Year Effect		
Specification	School-Grade-Year f.e.	Past S.-G.-Y. f.e.	Past Individual f.e.
Fraction in Grade			
Male	0.777**	0.867**	0.906**
	(0.162)	(0.182)	(0.185)
Free School Meal	-0.465*	-0.712**	-0.660**
	(0.230)	(0.242)	(0.246)
Special Needs	-0.206	-0.290	-0.061
	(0.163)	(0.180)	(0.182)
White	Ref.	Ref.	Ref.
Chinese	5.128**	4.436**	4.577**
	(1.645)	(1.689)	(1.712)
Mixed	0.222	0.143	0.239
	(0.526)	(0.553)	(0.562)
Indian	1.965*	1.866*	1.816*
	(0.787)	(0.816)	(0.832)
Bangladeshi	0.223	0.387	0.463
	(1.292)	(1.275)	(1.341)
Black African	0.933	0.935	0.850
	(0.744)	(0.755)	(0.762)
Pakistani	1.314	1.242	1.347
	(0.711)	(0.717)	(0.726)
Black, Other	-0.020	-0.173	-0.390
	(0.958)	(1.002)	(1.011)
Black Carribean	-3.030**	-3.195**	-3.192**
	(0.781)	(0.798)	(0.808)
School Fixed Effects	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes
R Squared	0.92	0.62	0.69
F Statistic	18,246.64	845.84	2,250.07
Number of observations	8,660,468	8,660,468	8,660,468
Number of school-grade-years	96,154	96,154	96,154
Number of schools	20,705	20,705	20,705

Source: National Pupils Database, Department for Education and Skills.

**: Significant at 1%. *: Significant at 5%.

Reading: Test Scores have a standard deviation of 10 and a mean of 50.

Table 10: Peer Effects in Schools in Grade 6 - Analysis of School-Grade-Year Effects

Specification	Dependent variable: ψ_j School Effect			
	School Effects	School-Grade-Year f.e.	Past S.-G.-Y. f.e.	Past Individual f.e.
School Status				
Community	Ref.	Ref.	Ref.	Ref.
Voluntary Aided	0.002	-0.031	0.049	0.061
	(0.047)	(0.054)	(0.061)	(0.060)
Voluntary Controlled	-0.430**	-0.372**	-0.215**	-0.194**
	(0.055)	(0.062)	(0.071)	(0.070)
Foundation	0.093	0.084	0.074	0.089
	(0.149)	(0.169)	(0.193)	(0.191)
Community Special	-3.367**	-4.875**	0.876**	0.876**
	(0.104)	(0.117)	(0.133)	(0.132)
Non-Maintained Special	-2.496*	3.659**	4.679**	4.814**
	(0.993)	(1.131)	(1.292)	(1.278)
Foundation Special	-4.442**	-6.571**	-2.778**	-2.737**
	(0.769)	(0.876)	(1.001)	(0.990)
R Squared	0.07	0.10	0.00	0.00
F Statistic	159.83	229.87	10.46	10.02
Number of observations	17,482	17,482	17,482	17,482

Source: National Pupils Database, Department for Education and Skills.

**: Significant at 1%. *: Significant at 5%.

Reading: Test Scores have a standard deviation of 10 and a mean of 50.

Table 11: The Analysis of School Effects – Grade 2

	Dependent variable: ψ_j School Effect			
Specification	School Effects	School-Grade-Year f.e.	Past S.-G.-Y. f.e.	Past Individual f.e.
School Status				
Community	Ref.	Ref.	Ref.	Ref.
Voluntary Aided	0.277**	0.426**	0.398**	0.368**
	(0.052)	(0.056)	(0.068)	(0.069)
Voluntary Controlled	-0.270**	-0.293**	-0.230**	-0.248**
	(0.061)	(0.067)	(0.080)	(0.081)
Foundation	0.201	0.146	0.012	-0.021
	(0.149)	(0.162)	(0.195)	(0.197)
Community Special	-3.404**	-4.315**	-1.643**	-0.877**
	(0.102)	(0.112)	(0.134)	(0.136)
Non-Maintained Special	-5.062**	-4.773**	-4.651**	-3.931**
	(0.460)	(0.502)	(0.605)	(0.612)
Foundation Special	-4.686**	-5.312**	-3.547**	-2.870**
	(0.822)	(0.897)	(1.080)	(1.093)
R Squared	0.08	0.10	0.02	0.01
F Statistic	182.45	228.14	36.34	18.17
Number of observations	16,197	16,197	16,197	16,197

Source: National Pupils Database, Department for Education and Skills.

**: Significant at 1%. *: Significant at 5%.

Reading: Test Scores have a standard deviation of 10 and a mean of 50.

Table 12: The Analysis of School Effects – Grade 6

λ	$\hat{\sigma}$	χ^2 Statistic	p-value
0.0	Ref.	Ref.	Ref.
0.1	+0.022	<0.000	>0.999
0.2	+0.022	<0.000	>0.999
0.3	+0.022	<0.000	>0.999
0.4	+0.022	<0.000	>0.999
0.5	+0.022	<0.000	>0.999
0.6	+0.022	<0.000	>0.999
0.7	+0.022	<0.000	>0.999
0.8	+51.700	5.014	0.025
0.9	+113.137	6.547	0.010

– 1998-2002 Cohort –

λ	$\hat{\sigma}$	χ^2 Statistic	p-value
0.0	+0.009	0.004	0.950
0.1	Ref.	Ref.	Ref.
0.2	+0.001	<0.000	0.984
0.3	+0.018	0.008	0.929
0.4	+0.014	0.006	0.938
0.5	+0.014	0.006	0.938
0.6	+0.009	0.004	0.950
0.7	+0.024	0.010	0.919
0.8	+3.777	3.263	0.071
0.9	+61.760	5.304	0.021

– 1999-2003 Cohort –

λ	$\hat{\sigma}$	χ^2 Statistic	p-value
0.0	Ref.	Ref.	Ref.
0.1	+0.001	<0.000	0.983
0.2	+0.009	<0.000	0.984
0.3	+0.010	<0.000	0.983
0.4	+0.001	<0.000	0.983
0.5	+0.001	<0.000	0.983
0.6	+0.002	<0.000	0.983
0.7	+0.029	0.013	0.909
0.8	+0.134	0.060	0.806
0.9	+7.948	2.074	0.150

– 2000-2004 Cohort –

Estimated with a2group, a2reg, xtlreg and xtlreg2. Available through the corresponding author Amine Ouazad.

See appendix B for the estimation technique for λ.

Source: National Pupils Database, Department for Education and Skills.

Under the hypothesis that λ is equal to its optimal value λ^* (0 for 1998-2002 and 2000-2004, 0.1 for 1999-2003), the statistic $2 \cdot ln(L(\lambda^*)/L(\lambda))$ converges to a χ^2 statistic. The statistic is reported in column 3 and the p-value of the test in column 4.

Table 13: Past and Current School-Grade-Year Effects - Residual Sum of Squares and the Discounting Factor λ

λ	$\hat{\sigma}$	χ^2 Statistic	p-value
0.0	Ref.	Ref.	Ref.
0.1	+750118.678	24.076	<0.000
0.2	+1,839,323.563	25.860	<0.000
0.3	+3,143,219.042	26.942	<0.000
0.4	+4,572,608.331	27.691	<0.000
0.5	+6,064,575.918	28.256	<0.000
0.6	+7,575,442.735	28.701	<0.000
0.7	+9,075,425.541	29.063	<0.000
0.8	+10,544,684.720	29.363	<0.000
0.9	+11,970,444.490	29.616	<0.000

– 1998-2002 Cohort –

λ	$\hat{\sigma}$	χ^2 Statistic	p-value
0.0	+261,013.403	21.860	<0.000
0.1	Ref.	Ref.	Ref.
0.2	+323,750.729	22.290	<0.000
0.3	+1,052,877.902	24.650	<0.000
0.4	+2,055,042.302	25.987	<0.000
0.5	+3,233,857.100	26.890	<0.000
0.6	+4,519,817.454	27.524	<0.000
0.7	+5,863,248.922	28.082	<0.000
0.8	+7,228,982.822	33.107	<0.000
0.9	+8,592,403.857	28.850	<0.000

– 1999-2003 Cohort –

λ	$\hat{\sigma}$	χ^2 Statistic	p-value
0.0	Ref.	Ref.	Ref.
0.1	+214,661.457	21.601	<0.000
0.2	+917,322.339	24.372	<0.000
0.3	+1,948,035.815	25.879	<0.000
0.4	+3,190,195.853	26.867	<0.000
0.5	+4,559,968.804	27.580	<0.000
0.6	+5,997,791.130	28.130	<0.000
0.7	+7,461,839.450	28.566	<0.000
0.8	+8,923,139.887	28.923	<0.000
0.9	+10,361,976.870	29.222	<0.000

– 2000-2004 Cohort –

Estimated with a2group, a2reg, xtlreg and xtlreg2. Available through the corresponding author Amine Ouazad.

See appendix B for the estimation technique for λ.

Source: National Pupils Database, Department for Education and Skills.

Under the hypothesis that λ is equal to its optimal value λ^* (0 for 1998-2002 and 2000-2004, 0.1 for 1999-2003), the statistic $2 \cdot ln(L(\lambda^*)/L(\lambda))$ converges to a χ^2 statistic. The statistic is reported in column 3 and the p-value of the test in column 4.

Table 14: Past and Current School-Grade-Year and Individual Effects - Residual Sum of Squares and the Discounting Factor λ

	Dependent variable			
	Mover	Moving to the most frequent school	School-Grade-Year effects	School effects
Past School-Grade-Year Effect	-0.003**	0.001**	0.563**	0.007**
	(0.000)	(0.000)	(0.002)	(0.001)
Past School Effect	0.020**	0.018**	-0.214**	0.341**
	(0.000)	(0.000)	(0.002)	(0.002)
Past Pupil Effect	-0.003**	0.001**	-0.051**	-0.023**
	(0.000)	(0.000)	(0.000)	(0.000)
Male	0.000	0.001	0.034**	-0.007*
	(0.001)	(0.001)	(0.004)	(0.003)
Free School Meal	0.032**	-0.031**	-0.148**	0.074**
	(0.001)	(0.001)	(0.006)	(0.005)
Special Needs	-0.015**	-0.002	-0.613**	-0.319**
	(0.001)	(0.001)	(0.006)	(0.005)
Month Of Birth	0.000	0.001**	-0.005**	-0.004**
	(0.000)	(0.000)	(0.001)	(0.000)
Chinese	0.014	-0.026**	0.492**	0.236**
	(0.007)	(0.006)	(0.037)	(0.031)
Mixed	0.007*	-0.027**	0.320**	0.281**
	(0.003)	(0.002)	(0.016)	(0.013)
Indian	0.017**	0.039**	0.529**	0.009
	(0.003)	(0.002)	(0.014)	(0.011)
Bangladeshi	-0.114**	-0.047**	1.509**	1.213**
	(0.004)	(0.003)	(0.027)	(0.022)
Black African	-0.038**	-0.087**	0.823**	0.939**
	(0.004)	(0.003)	(0.022)	(0.018)
Pakistani	-0.068**	-0.010**	0.473**	0.136**
	(0.002)	(0.002)	(0.014)	(0.012)
Black, Other	0.011*	-0.037**	0.295**	0.548**
	(0.006)	(0.005)	(0.031)	(0.026)
Black Carribean	-0.054**	-0.057**	0.390**	0.803**
	(0.003)	(0.003)	(0.019)	(0.016)
Number of observations	3,335,640	3,335,640	3,335,640	3,335,640
R Squared	0.01	0.01	0.52	0.16
F Statistic	1,019.03	1,289.69	71,912.88	8,117.50

Source: National Pupils Database, Department for Education and Skills.

**: Significant at 1%. *: Significant at 5%.

Reading: Test Scores have a standard deviation of 10 and a mean of 50.

Table 15: An Analysis of Mobility and the Direction of Mobility– Between Grade 2 and Grade 6

Number of Pupils	1,783,255	pupils	(100.00 %)
... with 2 years of observation	1,674,286	pupils	(93.89 %)
... changing school	750,022	pupils	(42.06 %)

			Percentage of moving pupils
Compulsory move	546,800	pupils	(72.90 %)
Non-compulsory move	203,222	pupils	(27.10 %)
Changing School Type	199,868	pupils	(26.65 %)
Changing LEA	112,592	pupils	(15.01 %)
Moving to the most frequent school	473,231	pupils	(63.10 %)
... among compulsory movers	423,995	pupils	(56.53 %)
... among non-compulsory movers	49,236	pupils	(6.56 %)

Compulsory movers: the pupil had to move, for his Key Stage 1 does not cater for Key Stage 2 pupils. Noncompulsory movers: the pupil could have staid in the same school for both Key Stage 1 and 2.
Source: National Pupils Database, Department for Education and Skills.

Table 16: Descriptive Statistics on Mobility

	Cannot stay in the same school	Can Stay in the Same School	
		Movers	Stayers
Moving to the most frequent school	0.776	0.151	-
	(0.417)	(0.358)	-
Male	0.508	0.509	0.507
	(0.500)	(0.500)	(0.500)
Month Of Birth	6.532	6.546	6.481
	(3.557)	(3.574)	(3.589)
Special Needs	0.213	0.267	0.214
	(0.409)	(0.442)	(0.410)
Free School Meal	0.156	0.244	0.156
	(0.363)	(0.429)	(0.362)
English spoken at home	0.928	0.920	0.915
	(0.259)	(0.272)	(0.279)
White	0.868	0.837	0.856
	(0.338)	(0.370)	(0.352)
Black Carribean	0.010	0.017	0.014
	(0.101)	(0.131)	(0.118)
Black, Other	0.004	0.007	0.004
	(0.064)	(0.085)	(0.067)
Pakistani	0.022	0.021	0.026
	(0.146)	(0.144)	(0.160)
Black African	0.008	0.019	0.012
	(0.087)	(0.137)	(0.107)
Mixed	0.017	0.024	0.018
	(0.128)	(0.154)	(0.133)
Bangladeshi	0.007	0.008	0.010
	(0.084)	(0.089)	(0.099)
Indian	0.023	0.018	0.020
	(0.151)	(0.132)	(0.141)
Chinese	0.003	0.003	0.003
	(0.050)	(0.057)	(0.051)

Compulsory movers: the pupil had to move, for his Key Stage 1 does not cater for Key Stage 2 pupils. Noncompulsory movers: the pupil could have staid in the same school for both Key Stage 1 and 2.
Source: National Pupils Database, Department for Education and Skills.

Table 17: The characteristics of compulsory movers, noncompulsory movers and nonmovers

	Dependent variable		
	Moving to the most frequent school	School-Grade-Year effects	School effects
Past School-Grade-Year Effect	0.002**	0.357**	0.009**
	(0.000)	(0.002)	(0.002)
Past School Effect	-0.002**	-0.178**	0.176**
	(0.000)	(0.003)	(0.002)
Past Pupil Effect	0.004**	-0.027**	-0.013**
	(0.000)	(0.000)	(0.000)
Male	0.000	0.021**	-0.003
	(0.001)	(0.006)	(0.005)
Free School Meal	-0.070**	-0.008	0.138**
	(0.002)	(0.010)	(0.008)
Special Needs	-0.003	-0.357**	-0.195**
	(0.002)	(0.009)	(0.008)
Month Of Birth	0.000**	-0.000	-0.001
	(0.000)	(0.001)	(0.001)
Chinese	-0.038**	0.500**	0.239**
	(0.012)	(0.059)	(0.049)
Mixed	-0.038**	0.298**	0.217**
	(0.005)	(0.025)	(0.021)
Indian	0.043**	0.754**	0.217**
	(0.003)	(0.020)	(0.016)
Bangladeshi	0.059**	1.389**	1.120**
	(0.007)	(0.051)	(0.042)
Black African	-0.040**	1.061**	1.091**
	(0.007)	(0.042)	(0.036)
Pakistani	0.057**	0.912**	0.486**
	(0.004)	(0.025)	(0.021)
Black, Other	-0.046**	0.274**	0.454**
	(0.010)	(0.053)	(0.045)
Black Carribean	-0.008	0.502**	0.747**
	(0.006)	(0.035)	(0.030)
Number of observations	1,088,517	1,088,517	1,088,517
R Squared	0.01	0.51	0.06
F Statistic	354.06	30,877.63	1,494.92

Compulsory movers: the pupil had to move, for his Key Stage 1 does not cater for Key Stage 2 pupils. Noncompulsory movers: the pupil could have staid in the same school for both Key Stage 1 and 2.
Source: National Pupils Database, Department for Education and Skills.
**: Significant at 1%. *: Significant at 5%.
Reading: Test Scores have a standard deviation of 10 and a mean of 50.

Table 18: An Analysis of Mobility and the Direction of Mobility for Compulsory Movers – Between Grade 2 and Grade 6

– School and Pupil Fixed Effects –

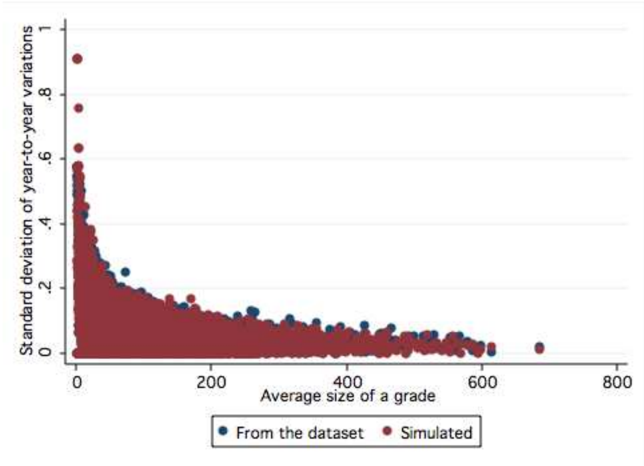
	Mean	Std. Dev.	y	θ	θ^*	$\theta^\perp$	ψ
y, Standardized grade	50.066	9.773	1.000				
θ, Pupil Effect	-0.000	9.191	0.812	1.000			
θ^*, Observed Pupil Heterogeneity	-0.000	6.252	0.475	0.680	1.000		
$\theta^\perp$, Unobserved Pupil Heterogeneity	-0.000	6.737	0.667	0.733	-0.000	1.000	
ψ, School Effect	8.241	6.938	0.074	-0.034	-0.017	-0.031	1.000
ε, Residual	0.000	4.364	0.447	-0.000	-0.000	0.000	0.000

– School Grade Year Effect and Pupil Effect –

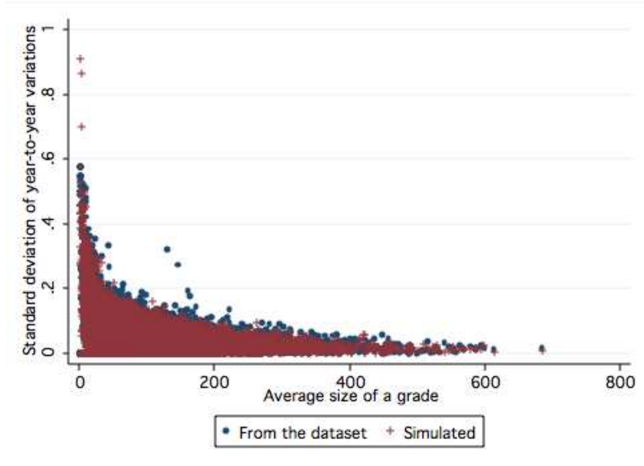
	Mean	Std. Dev.	y	θ	θ^*	$\theta^\perp$	φ	ψ
y, Standardized grade	50.066	9.773	1.000					
θ, Pupil Effect	-0.000	9.360	0.791	1.000				
θ^*, Observed Pupil Heterogeneity	-0.000	6.326	0.468	0.676	1.000			
$\theta^\perp$, Unobserved Pupil Heterogeneity	-0.000	6.898	0.644	0.737	-0.000	1.000		
φ, School-Grade-Year Effect	13.424	5.875	0.113	0.014	0.132	-0.102	1.000	
ψ, School Effect	-0.000	2.145	0.124	-0.092	-0.004	-0.121	0.355	1.000
ε, Residual	-0.000	4.277	0.438	-0.000	-0.000	0.000	-0.000	0.000

Estimated with a2group, a2reg, xtlreg and xtlreg2. Available through the corresponding author Amine Ouazad.
Source: National Pupils Database, Department for Education and Skills.

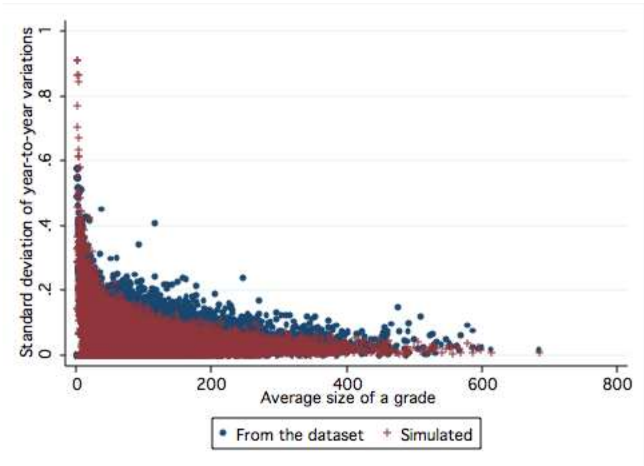
Table 19: Correlation Tables – Compulsory movers only



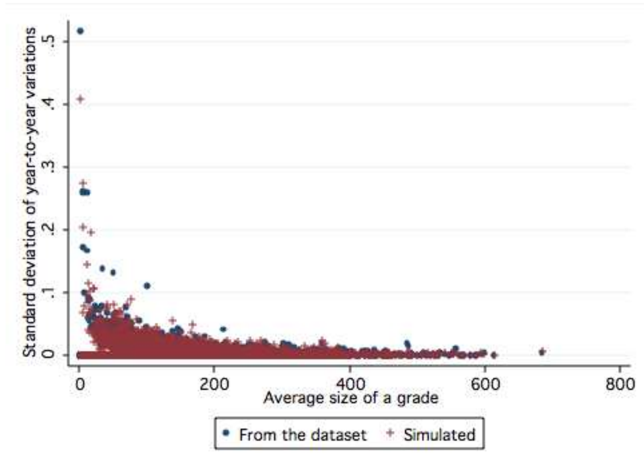
(a) Fraction Male



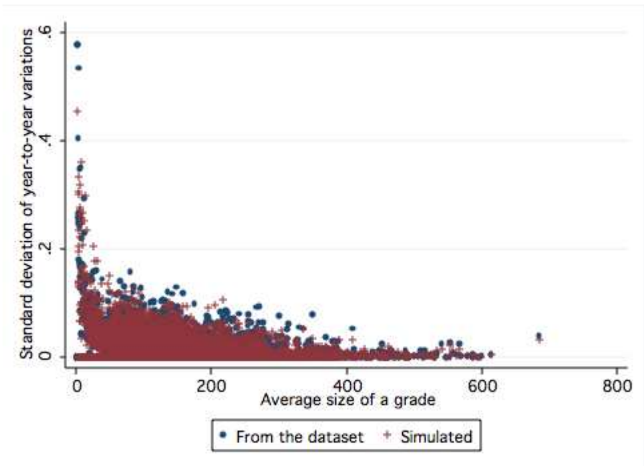
(b) Fraction on Free School Meals



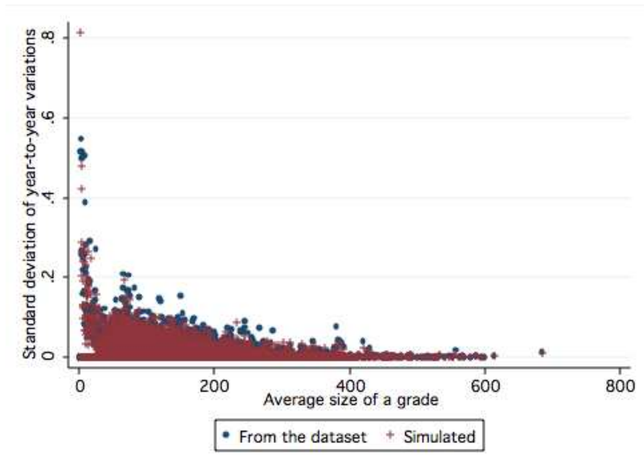
(c) Fraction of Special Needs



(d) Fraction Chinese



(e) Fraction Indian



(f) Fraction Black African

See section 5.3 for the description of the simulation procedure. Inspired by Lavy & Schlosser (2007).

Figure 2: Year to year variations in grade composition - Realized vs simulated deviations

	Overall Variance	Between Schools	Between LEAs	Between School Types
Key Stage 1				
Test Scores	99.006	16.172	1.094	3.475
	(100.0 %)	(16.3 %)	(1.1 %)	(3.5 %)
Individual Effects	76.607	17.278	1.755	2.922
	(100.0 %)	(22.6 %)	(2.3 %)	(3.8 %)
School-Grade-Year Effects	23.110	23.110	0.390	0.052
	(100.0 %)	(100.0 %)	(1.7 %)	(0.2 %)
School Effects	5.560	5.560	0.406	0.039
	(100.0 %)	(100.0 %)	(7.3 %)	(0.7 %)
Key Stage 2				
Test Scores	99.259	19.339	1.080	5.195
	(100.0 %)	(19.5 %)	(1.1 %)	(5.2 %)
Individual Effects	76.162	18.336	1.739	4.371
	(100.0 %)	(24.1 %)	(2.3 %)	(5.7 %)
School-Grade-Year Effects	13.924	13.924	0.390	0.057
	(100.0 %)	(100.0 %)	(2.8 %)	(0.4 %)
School Effects	5.160	5.160	0.406	0.044
	(100.0 %)	(100.0 %)	(7.9 %)	(0.9 %)

Source: National Pupils Database, Department for Education and Skills.

**: Significant at 1%. *: Significant at 5%.

Reading: Test Scores have a standard deviation of 10 and a mean of 50.

Table 20: Decomposition of Variance

– First Simulation –

	Mean	Std. Dev.	y	θ	φ
y, Standardized grade	49.893	9.958	1.000		
θ, Pupil Effect	0.000	9.271	-0.001	1.000	
φ, School-Grade-Year Effect	2.187	6.008	0.012	-0.033	1.000
ε, Residual	17.198	26.010	0.113	0.000	-0.000

– Second Simulation –

	Mean	Std. Dev.	y	θ	φ
y, Standardized grade	49.893	9.958	1.000		
θ, Pupil Effect	0.000	9.277	0.000	1.000	
φ, School-Grade-Year Effect	4.002	6.354	-0.011	-0.033	1.000
ε, Residual	17.201	26.010	0.113	-0.000	-0.000

– Third Simulation –

	Mean	Std. Dev.	y	θ	φ
y, Standardized grade	49.893	9.958	1.000		
θ, Pupil Effect	-0.000	9.271	0.001	1.000	
φ, School-Grade-Year Effect	-2.960	6.664	0.020	-0.032	1.000
ε, Residual	17.216	26.027	0.112	-0.000	0.000

Source: National Pupils Database, Department for Education and Skills.
Estimated with a2group, a2reg, xtlreg and xtlreg2. Available through the corresponding author Amine Ouazad.
Reading: Test Scores have a standard deviation of 10 and a mean of 50.

Table 21: Correlation Tables – Simulations