

Strand, S. (In press). **Do some schools narrow the gap?**
Differential school effectiveness revisited. *Review of*
Education (expected date of publication 2016).

I

Professor Steve Strand
Department of Education
University of Oxford
Oxford, England, UK
Tel: +44 (0)1865 611071
e-mail: steve.strand@education.ox.ac.uk

Keywords: equity, achievement gaps, school effectiveness, ethnicity, social class

Date of acceptance in *Review of Education*: 05/11/15

Journal homepage: [http://onlinelibrary.wiley.com/journal/10.1002/\(ISSN\)2049-6613](http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)2049-6613)

Word count 14,800

Note on contributor: Steve Strand is Professor of Education at the University of Oxford. His research interests are in equity issues in pupil achievement (particularly ethnicity, social class and gender) and school effectiveness and he has published widely in these areas.

Do some schools narrow the gap? Differential school effectiveness revisited

ABSTRACT

Relatively little research has explored whether schools differ in their effectiveness for different group of pupils (e.g. by ethnicity, poverty or gender), for different curriculum subjects (e.g. English, maths or science) or over time (different cohorts). This paper uses multilevel modelling to analyse the national test results at age 7 and age 11 of over 6,000 pupils attending 57 mainstream primary schools over three successive years in a socially and ethnically diverse inner London borough. The pupil groups with the poorest progress were White British pupils on Free School Meals (FSM) and Black Caribbean pupils, both those entitled and those not entitled to FSM. Differences between schools in average pupil progress were large, but there was no evidence of differential school effectiveness in relation to FSM, ethnicity or gender, all pupil groupings benefitted from attending the more effective schools and to a broadly similar extent. More effective schools 'raised the bar' but did not 'close the gap' suggesting that differences between schools in "quality" plays little role in equity gaps. While school residuals for different pupil groups were extremely highly correlated, school residuals were only moderately correlated across subjects (English, maths and science) and over time, with particularly poor stability for English.

INTRODUCTION

Concerns about unequal achievement in relation to ethnicity, socio-economic status (SES) and gender are long standing in England (e.g. Douglas, 1964; DES, 1985; Strand, 1999; Sirin, 2005) and recent topic reviews from the Department for Education (DFE) highlight the continued achievement gaps in contemporary data (see DfES, 2006; DCSF, 2007; and DCSF, 2009 for reviews in relation to ethnicity, gender and socio-economic disadvantage respectively). Successive governments in England have espoused the need to narrow or close such gaps, and the DFE has as one of its five strategic aims "closing the achievement gap between rich and poor" (DFE, 2014a, par 2.2). Much of the research has focussed on secondary school outcomes at age 16, but low achievement at the end of primary school at age 11 is a key risk factor for subsequent low achievement, for leaving full-time education at the earliest opportunity, and for long term employment and occupational outcomes (Boudon, 1974; McIntosh & Vignoles, 2000; Kingdon & Cassen, 2010). We know that socio-economic and ethnic achievement gaps exist at age 11 and that some have increased as pupils' progress through primary school. For example Strand (2010) analysed data for a national cohort of over 530,000 pupils in England tracked between age 7 and age 11 and showed a gap in progress between those pupils entitled to Free School Meal (FSM) and those not entitled to FSM of -0.13 SD (to -0.49 SD at age 11), and a Black Caribbean-White British gap in progress of -0.14 SD (to -0.28 SD at age 11). Poorer progress by low SES or FSM pupils is also reported age 4-7 (Strand, 1999) and age 11-16 (DfES, 2006; DCSF, 2009), although the White British-Black Caribbean progress gap age 11-16 seems specific to Black Caribbean boys from middle to high SES homes (Strand, 2014a).

One explanation for these ethnic and SES achievement gaps, and particularly for why they might grow over time, is that Black and low SES pupils attend schools of lower quality. For example in the US, Fryer and Levitt (2004) conclude that differences in 'school quality'

account for two-thirds of the growth in the Black-White gap between age 5 and age 7. In the UK, Wilson, Burgess and Briggs (2011) suggest that school quality accounts for around half of the Black Caribbean and Black Other gaps with White British, and Kingdon and Cassen (2010) also argue that ethnic minority pupils are more likely to attend poorer quality schools. However these conclusions are not supported by direct measures along traditional dimensions of school quality (such as average class size, teachers' qualifications, pupil-teacher ratios, school climate etc.) which do not correlate with the size of ethnic or SES gaps (e.g., Phillips et. al. 1998; Bali & Alvarez, 2004; Fryer & Levitt, 2006; Huang & Sebastian, 2015). Nevertheless, Government policy in England situates 'failing schools' as a significant part of the problem with an emphasis on minimum requirements for pupil achievement known as floor standards, and a forced programme of school conversion (academisation¹) for schools failing to meet the floor standards. Indeed by August 2014 there were no fewer than 1,112 sponsored academies, 93% of which had been formed from 'underperforming Local Authority maintained schools' (House of Commons, 2015, p7). The DFE (2014b, p10) has signalled a clear intention for this approach to continue with a rise from the 2015 age 11 floor standard of 65% of pupils achieving the expected standard to a minimum of 85% of pupils achieving the expected standard in all of reading, writing and mathematics in 2016. How strong though is the evidence that schools are central in the creation of achievement gaps?

Differential effectiveness by pupil groupings

It is well established that schools have different levels of effectiveness as indexed by Contextual Value Added (CVA) measures after adjustment for intake characteristics such as prior attainment and socio-economic factors (e.g. Rutter et al, 1979; Mortimore et. al. 1988; Strand, 1997). However if school effectiveness or quality was the main influence on SES or ethnic achievement gaps, then these gaps would be expected to be larger in some schools and smaller in others. For example some schools may be more effective in promoting the progress of low SES than high SES pupils, or boys versus girls, or some ethnic groups more than others. This is termed *differential* school effectiveness. These questions are of substantial importance since if schools do differ significantly in terms of their effectiveness for particular pupil groupings, then an investigation of the factors associated with differential effectiveness will be important for the design and implementation of policies on equal opportunities (Kyriakides, 2004).

There is very little research on the phenomenon of differential school effectiveness (Kyriakides, 2004, p143) and current evidence on the existence of differential school effects is mixed. There are some positive reports of differential effectiveness in relation to ethnicity and prior achievement from studies of secondary schools in England (e.g., Nuttall, Goldstein, Prosser & Rasbash, 1989; Thomas, Sammons, Mortimore & Smees, 1997b; Dearden, Mickelwright & Vignoles, 2011). However other studies describe negative findings, including those from primary schools (e.g., Brandsma & Knuver, 1989; Smith & Tomlinson, 1989; Sammons, Nuttall & Cuttance, 1993; Palardy, 2008; Strand, 1999, 2010, 2014b). For example in the study discussed earlier Strand (2010) directly tested for differential school effects on pupil progress age 7-11 in relation to prior attainment, ethnicity, FSM and gender using an English national dataset of 530,000 pupils attending over 14,200 primary schools. He reports statistically significant but extremely small differential effects in relation to prior attainment, FSM and gender, and no differential effects at all in relation to ethnic group. No school appeared to eliminate or reverse the typical within-school gap in relation to FSM or the Black Caribbean-White British gap. Additionally the correlation between schools' residuals for pupil entitled to FSM and those not entitled to FSM was 0.97, as was the correlation between

school residuals for White British and for Black Caribbean pupils. The correlation between schools' residuals for boys and for girls was 0.98. Strand concludes that the same schools that are most effective for White British pupils, boys or pupils on FSM are also the most effective for Black Caribbean pupils, girls and those not on FSM. In short, there was no evidence of differential school effectiveness for different pupil groupings.

Strand (2010) did however express some caution in his conclusions, given the relatively small sample size in many primary schools (average $n=37$ pupils) and the resulting shrinkage towards the mean in multi-level estimates, particularly for small pupil groups within schools. The present study will address this issue by focussing on a large inner London Local Authority serving an ethnically and socially diverse population, and collecting data over three successive cohorts to substantially boost within school sample sizes. Data on national tests for 11 year olds in English, mathematics and science were collected in 2004, 2005 and 2006, along with prior attainment scores at age 7. This yielded a total sample of over 6,000 pupils across 57 schools with an average within school sample of 120 pupils per school.

Differential effectiveness across subjects or over time

The availability in this study of data from standardised tests in different subjects and over three successive years also allows the possibility of assessing other differential effects, particularly whether schools vary significantly in their effectiveness in relation to different subjects (English, maths and science) and the extent to which schools that are effective in any one year are equally effective in subsequent years? Research in secondary schools suggests quite marked differences between Contextual Value Added (CVA) scores for subject departments within schools. For example Thomas, Sammons, Mortimore and Smees (1997a) analysed pupil progress between age 11 and a variety of GCSE subject examinations at age 16 in a sample of 69 inner London secondary schools. While the correlation between school VA measures for total GCSE score and English ($r=0.57$) and total GCSE score and mathematics ($r=0.48$) were moderately high, the direct correlation between schools' VA scores for English and maths was only $r=0.24$. Overall seven out of 15 inter-subject correlations reported were less than $r=0.30$, indicating that even within 'effective' schools there are less effective departments and *vice versa*. There is also evidence of secondary schools varying in their effectiveness over time, with moderate to high correlations between CVA estimates in adjacent years but little evidence of past performance being particularly predictive over longer periods of up to five years (Thomas, Sammons, Mortimore & Smees, 1997a; Gray, Goldstein & Thomas, 2001; Manahan, Pugh & Gray, 2005; Leckie & Goldstein, 2011). Results vary though according to the particular method used to estimating school performance, for example three year rolling averages are more stable than year on year comparisons (e.g., Thomas, 2001), and particularly low estimates have been reported for growth curve models (e.g. Dumay, Coe & Anumendum, 2014).

For primary schools there is much less research in relation to these questions, but results suggest correlations of schools' VA/CVA estimates across subjects such as reading and mathematics tend to be higher than for secondary schools, typically in excess of $r=0.60$ (e.g. Luyten, 1998; Sammons et al, 1993; Hill & Rowe, 1996; Strand, 1997). Additionally Ma (2000) reports that the size of socio-economic gap within schools tend to be highly correlated across subjects (reading, writing, maths and science) and this seems consistent in both raw and VA/CVA models. Very little research has looked at stability over time for primary schools. Day, Stobart, Sammons, Kington, Gu, Smees & Mujtaba (2006) compared the stability of teachers CVA scores for maths calculated over two successive years and report moderate correlations

of $r=0.48$ for Y6 teachers and $r=0.51$ for Y9 teachers, though sample sizes were small ($n=37$ and 38 respectively). Dumay et. al. (2014) analysed Y5 maths test scores for a sample of 10,000 pupils in 200 UK primary schools between 2000 to 2004 and report correlations for school VA measures (from a regression model adjusted for pupil prior attainment at the end of Y1) of $r=0.40$ to $r=0.53$ in adjacent years but dropping to $r=0.30$ for VA estimates five years apart. Perhaps the most comprehensive contemporary study is that of Melhuish, Romaniuk, Sammons, Sylva, Siraj-Blatchford & Taggart (2006) who analysed national tests for 11 years olds in English, maths and science over the three years 2002, 2003 and 2004 for all primary schools in England, together with pupils prior attainment at age 7. They report that school residuals from CVA models correlated only moderately between subjects, ranging from 0.56 between English and maths to 0.74 between maths and science. However most notable were the particularly low correlations over time, particularly over the three year period 2002 vs. 2004, with correlations for school CVA residuals of only 0.29 for English, 0.42 for maths and 0.45 for science. They conclude that in all subjects, but particularly in English, there is considerable instability in primary schools' levels of effectiveness over time. However they used relatively undifferentiated outcome measures (National Curriculum levels rather than test marks) and employed a series of separate univariate analyses rather than multivariate analyses which are more appropriate given that English, maths and science scores are positively correlated. Multivariate models allow for correlation between different dependent variables, offering greater statistical power and considerably reducing the risk of both type 1 and type 2 errors (see De Mayer et al, 2010 for a discussion).

Plan of the paper

After presenting some simple descriptive statistics, the analysis first explores the overall (or fixed) effects of pupil background variables on progress age 7-11 for each subject. Next the paper considers whether pupil progress does indeed vary significantly between schools and the size of the 'school effect' for each subject. Having established that large differences between schools exist, the paper then considers the question of differential school effectiveness by subject, over time and for different pupil groupings respectively. The results conclude with an analysis of the correlations between schools' 'raw' results and their results from both Value Added (VA) and Contextual Value Added (CVA) models of pupil progress. The implications of the results for understanding the factors underlying the growth in equity gaps over time are considered. Finally some wider implications for the presentation of data in school performance tables are discussed.

METHOD

The Local Authority

The Local Authority (LA) is a densely populated inner London borough serving one of the most ethnically and linguistically diverse populations in England. Almost 40% of pupils speak a language in addition to English with over 150 languages spoken at home. Four-fifths (80%) of pupils are from ethnic minority groups, with the three largest ethnic groups being Black Caribbean (23%), Black African (21%) and White British (20%). The LA has high levels of deprivation with one-third (38%) of pupils entitled to FSM, nearly 2.5 times the national average for primary schools of 16% (DCSF, 2009). The LA was ranked the 23rd most deprived of the 354 Local Authority districts in England on the Government Index of Multiple Deprivation 2004 (IMD, 2004, Annex L). The dataset analysed here consists of the 2004, 2005 and 2006 national test results of a total of 6,882 pupils aged 11 drawn from all 57 maintained primary schools in the LAⁱⁱ.

Dependent variables

Age 11 test scores: At age 11 pupils in England complete national tests in English, mathematics and scienceⁱⁱⁱ. New tests are set each year and are externally marked. Outcomes are often reported as National Curriculum (NC) levels, with the 'typical' pupil expected to achieve level 4 and higher attaining pupils achieving level 5. However as outcome variables NC levels lose the more detailed information contained in the raw test marks. This study used as dependent variables pupils marks in English (reading, writing and spelling tests, range 0-100 marks); mathematics (mental maths test and two written tests one allowing and one not allowing the use of calculators, range 0-100 marks); and science (two written tests, range 0-80 marks). In addition an average age 11 score was calculated across all tests (range 0-280 marks). Because the national tests change every year it is not possible to assume that scores in different years are strictly comparable. Therefore all marks were normal score transformed each year to give a mean of zero and SD of 1 for each outcome in 2004, 2005 and 2006 respectively.

Explanatory variables

Age 7 average test score: Prior attainment was assessed through national tests in reading, writing and mathematics completed at age 7. Tests are set each year and are marked by teachers within schools. Outcomes are reported as NC points scores ranging through Working towards level 1, level 1, level 2c, level 2b, level 2a and level 3 for higher attaining pupils. Pupils' points scores were averaged across the three tests and transformed to a normal score with mean of zero and SD of 1. The rationale for using the average age 7 score rather than the separate components was threefold. First, and most importantly, the average score at age 7 was the best single predictor for all outcomes at age 11 with correlations of 0.74, 0.71 and 0.66 with English, maths and science respectively. Second, the average age 7 score provided a normally distributed variable spread over 23 values compared to just six values within each age 7 test component. Third, an overall measure of prior attainment facilitates the analysis of differential effectiveness by providing a single term that can be interacted with other variables.

Entitlement to a Free School Meal (FSM): Pupils from families in receipt of state benefits such as income support, jobseekers allowance or child tax credits (where family income is below 60% of the national median) are eligible for FSM, making it a direct measure of family poverty. While a more differentiated measure of family socio-economic status (SES) would be preferable, entitlement to a FSM is the only pupil level SES variable routinely collected in England pupil datasets.

Ethnic group: Pupil's ethnic group was recorded in 21 extended ethnic categories. The largest groups were Black Caribbean (23%), Black African (21%) and White British (20%), together accounting for nearly two-thirds of all pupils. The number of pupils in the smaller ethnic categories were often low, so to form sufficiently large groups for analytic purposes some of the smaller groups were combined to give a total of ten ethnic groups with a minimum of 150 pupils per group^{iv}. Although White British are not the largest group, for the purpose of comparison they were used as the reference group since they are the majority group in England.

Gender: Boys are coded 0 and girls coded 1.

Age within year group: In England pupils are organised into year groups based on their date of birth. Pupils within a year group can therefore vary in age by almost 12 months with the oldest pupil being born on the 1 September at the start of the academic year and the youngest born on 31 August the following calendar year. Pupil age in completed months at the time of sitting the age 11 tests was calculated. This value was deviated from the sample average (134 months) to give a variable with a mean of 0 and range from -5 (for the youngest pupils) up to 6 (for the oldest pupils).

Special educational needs (SEN): Pupils without any identified SEN were the reference group with three dummy variables for different stages of identified SEN. The first stage (*School Action*) is where the school identifies the need and makes provision from its own resources. The second stage (*School Action Plus*) requires the school to involve an external professional (often an Educational Psychologist) in identifying and planning for the pupil's need. The final stage, for pupils with the greatest needs, is a statutory assessment by the Local Authority which may lead to a formal *statement of SEN* and additional external resources to meet the need.

Pupil mobility: Pupils who remained in the same school throughout the age 7-11 period were the reference group with four dummy variables for those who joined their school in Y3, Y4, Y5 and Y6 respectively.

Stage of Fluency in English: For pupils with English as an Additional Language (EAL) teachers rated the pupils' degree of fluency in the English language on a four point scale ranging from complete beginner (stage 1) through to fully fluent in English (stage 4). Mono-lingual English speakers were coded 0. Judgements were moderated across the LA by a team of peripatetic teachers.

School composition: For each cohort within each school, composition variables were created by aggregating the relevant pupil characteristic, e.g. the % of pupils entitled to FSM; the % of girls, the % pupils from each ethnic group and so on. There were a total of 171 observations (57 schools * 3 years) for each composition variable.

Sample selection

Listwise deletion of a small number of cases with missing values on one or more of the pupil background characteristics gave a total sample of 6,882 Year 6 pupils. The majority of these (6045 or 87.8%) also had prior attainment (age 7) scores although a minority (12.2%) had entered their schools from outside England sometime after age 7. This is typical of London, the capital city, which experiences high levels of international migration. For these pupils it is not possible to complete an analysis of pupil progress. Since the main purpose of this study was to explore variations in pupil progress and school effects on progress, it was considered important to have a balanced sample across all analytic models (raw, value added and contextual value added models). Only pupils with age 7 scores were therefore included, giving a final sample of 6,045 pupils.

RESULTS

Descriptive statistics

Table 1 presents descriptive statistics for the sample including for each value of the pupil background variables the number of observations, the percentage this represents among the sample, the average age 7 score and the average age 11 English, mathematics and science scores. These data can be analysed to reveal the extent of various achievement gaps. For example at age 7, pupils entitled to FSM scored $-.29$ SD below the mean while those not on FSM scored $.17$ SD above the mean, a FSM gap of $-.47$ SD. Similarly we can compute the gaps between each ethnic group and White British pupils. For example at age 7 Black Caribbean pupils scored $-.15$ SD below the mean while White British pupils scored of $.20$ SD above the mean, a gap of $-.35$ SD. The gender gap at age 7 (boys $-.13$ SD vs. girls $.11$ SD) was smaller at 0.24 SD in favour of girls.

< ----- >
Insert Table 1 about here
< ----- >

Some of the above gaps do not change between age 7 and age 11. For example as illustrated above the FSM gap was $-.47$ SD at age 7, and this gap was substantially the same at age 11; $-.45$ SD for English, $-.47$ SD for maths and $-.46$ SD for science. However the data indicate that the gaps for some ethnic minority groups narrow, and for other groups they grow, between age 7 and age 11. For example the gap between Portuguese and White British pupils was $-.86$ SD at age 7, but narrowed to $-.57$ SD for English, $-.55$ SD for science and $-.42$ SD for maths at age 11. Conversely the gap between Black Caribbean and White British pupils increased from $-.35$ SD at age 7 to -0.42 SD for English and $-.50$ SD for both maths and science at age 11. The next section will explicitly model variation in progress age 7-11 in relation to pupil background characteristics. The regression analysis reported will allow us to account for confounding between ethnicity and FSM, for example 32% of White British pupils are entitled to FSM, compared to 39% of Black Caribbean and 45% of Black African pupils. It will also allow for interactions between ethnicity and FSM, given these have been frequently identified in previous research (e.g. Strand 1999, 2010, 2014a).

Regression Modelling

In order to determine the unique influence of each variable on pupil progress age 7-11 we turn to multiple regression analysis. School effects were a key interest here, and to capture the hierarchical nature of the data multi-level regression models were used. A multivariate multi-level analysis was conducted to simultaneously model effects on English, mathematics and science outcomes. Analyses were completed using the MLWin package (V2.25) with age 11 test scores (level 1) grouped within pupils (level 2) grouped within cohorts (level 3) grouped within schools (level 4). Level 1 exists only to model the multivariate nature of the data, with pupils, cohorts and schools as the substantial hierarchical levels. In the multivariate analysis separate coefficients were specified for each outcome (English, maths and science). For comparison purposes a univariate analysis was also completed on average age 11 test score. The results of the fixed part of the models are presented in Table 2^v.

< ----- >
Insert Table 2 about here
< ----- >

Fixed effects

Pupil characteristics

For ease of presentation we shall focus here on the association with average age 11 test score, but noting any significant variations that occur for individual subjects. The effects for gender, pupil age within year group, level of SEN, mobility and stage of fluency in English were straightforward to interpret. Girls made slightly less progress than boys overall (-.06 SD) although this reflected girls making more progress than boys in English (0.17 SD), but less progress in maths (-.21 SD) and science (-.11 SD). Older pupils made slightly less progress than their younger peers with a -.05 SD contrast between pupils three months below/above the mean age. Pupils identified with SEN made significantly less progress than pupils without identified SEN (-0.26, -0.38 and -0.58 SD for school action, school action plus and statemented pupils respectively). The effect of mobility on pupil progress depends upon when the pupil joined their new school, with a significant negative impact on progress only for those who joined in Y6. Pupils identified as beginners or needing considerable support in learning English made significantly less progress (-.26 SD) than their monolingual peers, while those identified as fully fluent made slightly more progress (0.06 SD), although the contrasts were slightly lower in maths than the other two subjects.

However the results for ethnic group and FSM need to be interpreted in the context of a significant interaction between these two variables for all outcomes^{vi}. It is not possible therefore to talk about the main effect of ethnic group or the main effect of FSM, these can only be understood in the context of their interaction. To illustrate this Figure 1 presents progress age 7-11 separately for FSM and Non-FSM pupils in the three largest ethnic groups (Black African, Black Caribbean and White British) for each subject and for average age 11 test score^{vii}. In interpreting the figures the central horizontal 'zero' line indicates average progress 7-11 across the entire sample. Foregrounding FSM, the FSM gap in progress for average age 11 score for White British pupils was -.25 SD, which was substantially larger than the FSM gap for Black African pupils (-.06 SD) and for Black Caribbean pupils (-.04 SD). Foregrounding ethnicity, among those not entitled to FSM White British pupils made the greatest progress by a substantial margin (although their advantage over Black African pupils was not significant for English). In comparison, among those on FSM there were only small differences between ethnic groups, with no significant difference in progress for average score between White British and Black Caribbean pupils (although there was a significant contrast for maths) and with Black African pupils actually scoring slightly higher (driven by a particular strong performance in English)^{viii}.

In summary as has been noted previously (Strand 2010, 2012) it is White British pupils not on FSM that make the greatest progress age 7-11, widening the FSM gap over time and doing substantially better than their ethnic minority peers. In contrast poor progress age 7-11 is equally notable for both White British and Black Caribbean pupils entitled to FSM, although Black African pupils as well as some other ethnic minorities, particularly Bangladeshi/Pakistani, Portuguese and Other Asian, buck this trend (see Table 2).

< ----- >
Insert Figure 1 about here
< ----- >

School composition

A range of school composition variables were also entered, these variables are described in Table 3. All the school composition variables were initially included in the model and then removed via backwards elimination. Two school composition variables had a statistically significant association with pupil progress in at least one subject. First, even after accounting for individual pupil FSM status, there was a significant negative association with the percentage of pupils in the cohort entitled to a FSM for all three outcomes. Specifically a one point increase in the percentage of pupils entitled to FSM was associated with a $-.006$ SD decrease in age 11 English score, with a similar impact for maths and science. Given the school mean %FSM had a SD of 15, this represents an effect size (ES) of -0.18^{ix} . This is consistent with much previous research suggesting schools with high levels of socio-economic deprivation face additional barriers to pupil progress (e.g. Strand, 1998, 2010; Caro, 2011). Second, there was also a negative association with progress for the cohort mean average age 7 score ($-.21$ SD, ES= -0.15 SD), although analysis of the separate subjects indicates this was driven by a large effect for mathematics ($-.27$ SD, ES= -0.18) with smaller and non-significant effects for English and science (both $-.13$ SD, ES= -0.10). The direction of the effect may initially appear surprising, more able cohorts on average making less progress, but is consistent with much previous research with national tests in England (e.g. Strand, 1997; 2010). The negative compositional effect may reflect the fact that age 7 tests are marked by the teacher, while the age 11 tests are externally marked. If some teachers are generous in their marking at age 7, these classes may do slightly less well in the externally marked age 11 tests, giving rise to the negative compositional effect. However it should be remembered that the ES for this age 7 composition effect is only significant for mathematics and even then is small compared to the ES for the individual pupil age 7 score of 1.22 SD.

< ----- >
Insert Table 3 about here
< ----- >

Random effects

We turn now to the random part of the model to explore the variation at each hierarchical level (pupils, cohorts and schools). The results are presented in Table 4. Three multivariate models are compared.

- A **Null model** with just an intercept term random at each level. This is an analysis of 'raw' scores and is a baseline against which subsequent models can be compared.
- A **Value Added (VA) model** which adds just the age 7 prior attainment scores (each pupils' age 7 score, age 7 score squared and cohort mean age 7 score).
- A **Contextual Value Added (CVA) model** including not just prior attainment but all the pupil background (age, FSM, gender, mobility, fluency in English, SEN, ethnicity) and school (%FSM and cohort mean age 7 score) variables as listed in Table 2.

Because this is a multivariate model there are separate results for each of the English, maths and science outcomes. In all models only the intercept term is random at each level.

< ----- >
Insert Table 4 about here
< ----- >

There are three key points to take from Table 4. First, the proportion of variance at the school level in the null model is around 12% for English and maths, though somewhat higher (17%) for science. A substantial part of this variation, particularly for English, can be explained by prior attainment at age 7, with a reduction in the school level variance (VR) in the VA model of

over 63% for English (around 42% for the other two subjects). The full CVA model explains an even greater proportion of the variance, accounting for 77% of the school level variance for English and 57% and 53% of the school level variance for maths and science respectively. The substantial increase in the variance reduction for the CVA model against the VA model attests to the importance of considering a range of pupil characteristics, other than just their prior attainment, in understanding variation in pupil progress between schools.

Second, there is still substantial variability between schools even after all the controls included in the CVA model, which may be interpreted as ‘school effects’ on pupil progress age 7-11 net of the measured pupil and school background characteristics. The school level variance in intercepts in the CVA models in Table 4 are 0.023, 0.049 and 0.074 for English, maths and science respectively. Taking the square root of these variances gives the standard deviation (SD) of school variability around the LA average, and an indication of the size of the school effect can taken as 2*SD, i.e. the difference in average pupil progress between the more effective schools (one SD above the LA average) and the less effective schools (one SD below the LA average). Estimates of school effects are thus 0.30 SD for English, 0.44 SD for maths, 0.54 SD for science and 0.52 SD for average age 11 score. A graphical display of the CVA scores for the 57 primary schools is given in Figure 2 together with 95% confidence intervals. Only three schools had English CVA scores that could be reliability distinguished from the LA average, reflecting the substantial year to year variation within schools for English (discussed below). However 17 schools had maths scores that differed significantly from the LA average and 20 schools had science scores that varied significantly from the LA average. For average age 11 score 15 schools had CVA estimates significantly different from the LA average.

< ----- >
 Insert Figure 2 about here
 < ----- >

Third, the year level in the model captures variation over time within schools, specifically the within-school cohort variation. There are important changes in this variation both across different subjects at age 11 and between null, VA and CVA models. Considering first differences between subjects, we see from the null model that there is considerably greater within-school cohort variation for English (5.6%) than for science (3.8%) or maths (2.0%). Another way of expressing this is by calculating the within-school cohort variance as a proportion of the total between school variance, or alternatively the average correlation between cohorts within schools. These results are presented in Table 4 in the line headed cohort estimates^x. There is much more within-school cohort variation for English (33%) than for maths (14.4%) or science (18.2%). Thus the correlation between cohorts within schools is much lower for English (0.67) compared to both science (0.82) and maths (0.86). This relative difference between subjects continues to be seen in the VA and CVA models with the proportion of within-school cohort variation for English consistently around 2.5 times greater than for maths.

Considering differences between models, it is apparent that year to year within-school variation is greater in the VA and CVA models than the null models. For example for English the proportion of variance at the year level increases from 5.6% in the null model to 14.9% in the VA model and 16.9% in the CVA model. This gives a negative variance reduction (VR) at the year level for the VA and CVA models, although this is just indicating that the within-school cohort variation increases in the VA and CVA models relative to the null model. We

can see this in the average correlation (r) between cohorts within schools, which for English reduces from 0.67 in the null model to 0.37 in the VA model and just 0.28 in the CVA model. Clearly the VA and CVA estimates for English are much less stable over time than raw results. The changes are nothing like as drastic for maths and science, but the pattern of lower correlations for VA and CVA models relative to the null model is the same. In short schools vary a lot more over time in their VA and CVA estimates than they do in their raw scores.

The above results suggest it will be valuable to directly calculate estimates of residuals for each cohort within each school (and for each subject). To do this we calculate a three level (subjects within pupils within schools) model where the overall intercept term is removed and the year variables (2004, 2005 and 2006) are allowed to vary randomly at both the pupil and school level. There are now no within-school cohort random effects because the school level covariance matrix completely explains the variation between school cohorts so there is no residual variation to explain. Given the response variables are normalised to have a mean of zero and SD of 1 in each year, this parameterisation of the model provides an identical fit to the data as the four level model but with the ability to directly estimate residuals for each school in each subject in each year (see Leckie, 2013, p36-37). These results are analysed below to examine differential school effectiveness by (i) curriculum subject, (ii) year (i.e. over time), (iii) pupil groupings (ethnicity, gender and FSM), and (iv) across models (null, VA and CVA).

(i) Differential school effectiveness by subject

At the *pupil level* there were high correlations between subject scores (average correlations English vs. maths= 0.75, English vs. science= 0.76, and maths vs. science= 0.81). Generally pupil who did well in one subject tended to do well in others. However the key interest here is in the *school level* correlations; do schools that perform well in one subject tend to perform well in others, and how does this vary across the raw, VA and CVA models? Table 5 presents the school level correlations between subjects for each model.

< ----- >
Insert Table 5 about here
< ----- >

School level correlations are highest for raw scores, ranging from 0.77 to 0.91. Correlations reduce somewhat in the VA model showing more variability across subjects, ranging from 0.61 between English and science in 2006 to 0.88 between maths and science in 2005. Correlations in the CVA model are only slightly attenuated compared to the VA model, ranging from 0.56 to 0.87. The final column of Table 5 presents the average between-subject correlation across years and indicates that, even between English and maths where the relationship is the weakest, the correlation between school CVA residuals for the two subjects is 0.65, and for maths and science it is much higher 0.83. As an example Figure 3 plots the school CVA residuals for English and maths in 2006 ($r=0.66$). We conclude that in schools where pupils make strong progress in one subject they also tend to make strong progress in the others, although the relationship is by no means perfect.

< ----- >
Insert Figure 3 about here
< ----- >

(ii) Differential school effectiveness over time

Table 6 presents the correlations between school residuals for each subject across years. The average correlation between schools' raw scores across years tend to be high for maths ($r=0.86$) and science ($r=0.82$), but markedly lower for English ($r=0.66$). However this does mask the finding that correlations only one year apart (i.e. 2004 vs. 2005 or 2005 vs. 2006) tend to be higher than the correlations two years apart (i.e. 2004 vs. 2006). Correlations between school VA residuals are lower than in the null model but still high for maths ($r=0.71$) and science ($r=0.68$), indicating that schools where pupils make strong progress in these subjects in any one year also tend to make good progress in other years. However the correlation for English are particularly low (average $r=0.38$) indicating low stability over time: schools with strong progress in English in one year may have quite different results in other years. The CVA correlations do not differ greatly from the VA correlations, so while additional controls for pupil and school characteristics in CVA models increase significantly the proportion of variance that can be explained it makes relatively little difference to the rank order of schools. Figure 4 presents the correlation between CVA mathematics scores in 2004 and 2006. While the correlation is moderate ($r=0.58$) we see that even for maths there is quite some variability in schools CVA scores two years apart.

< ----- >
Insert Table 6 and Figure 4 about here
< ----- >

The reason for the particularly low stability over time of English scores is unclear. Looking at the correlations between raw scores, it is noticeable that while correlations between years for English are always lower than the equivalent correlations for other subjects, the English inter-correlations involving 2006 results ($r=0.60$ and 0.63) are much lower than between 2004 and 2005 ($r=0.75$). There were no major changes to the structure of the tests over the three years. The specific test items do vary year to year and average marks in 2006 (mean=57) were higher than in the previous two years (mean=52), but this would have been accounted for by our within-year normalisation. The correlation between age 7 score and English score was somewhat lower in 2004 ($r=0.72$) than in 2006 ($r=0.76$), but the same was true for maths ($r=0.64$ in 2004 vs. $r=0.69$ in 2006). The finding appears robust since Melhuish et. al. (2006) report similar results based on national data for England 2002-2004, with average CVA correlations over time of 0.33 for English compared to 0.51 for maths and science (p40)^{xi}. It is possible that there is greater noise in the English results because the reliability of the English test is lower than for maths and science, with an analysis of 2009 and 2010 data reporting Cronbach's alpha of 0.91, 0.93 and 0.97 for English, science and maths and classification accuracy of 85%, 87% and 90% respectively (He, Hayes & Wiliam, 2013). In particular the reliability of the English writing test has always been contentious because of greater subjectivity, and indeed the statutory writing test at age 11 was removed in 2013. Alternatively, we know from previous research that schools/classes account for much less of the variation in English progress than they do for mathematics or science, because the later two subjects are predominantly learnt in schools whereas home influences are stronger for literacy (e.g. Luyten, 1998; Hill & Rowe, 1996). We have only a crude measure of family background in FSM, English fluency and SEN, and it may be that unmeasured factors related to home circumstances, particularly factors such as parental instruction, underlie the greater variability in CVA English scores.

(iii) Differential school effectiveness by pupil characteristics

A key question for this study was the extent to which ethnic, FSM or gender gaps in pupil progress varied across schools. However trying to incorporate random variation in pupil characteristics alongside year and subject variance within the multivariate model produced a very complex random structure at the school level with the covariance matrix expanding markedly. Therefore to test for differential effectiveness in relation to pupil characteristics separate univariate models were completed for English, mathematics, science and average age 11 score^{xii}. Year was incorporated as a hierarchical level with pupils (level 1) grouped within cohorts (level 2) grouped within schools (level 3) to allow for the substantial within-school cohort variance when estimating school effects. Tests were completed in the CVA models for school level variation in each pupil characteristic separately. The results are presented in Table 7.

< ----- >
Insert Table 7 about here
< ----- >

It is clear that for all three subjects as well as for average age 11 score the only significant school variation was in relation to pupil prior attainment at age 7. For the average age 11 score outcome, 15 schools had age 7 slopes that differed significantly from the LA average. This indicates that for a significant minority (26%) of schools their CVA estimates for pupils with low prior attainment (1 SD below the LA mean age 7 score) and pupils with high prior attainment (1 SD above the LA mean age 7 score) differed significantly. To visualise this Figure 5 plots the school CVA estimates for average age 11 score at three levels of prior attainment. While it is clear that for a small number of schools there were quite large differences in CVA at different levels of prior attainment, overall the school CVA estimates for pupils +/- 1 SD above/below the mean age 7 score correlated 0.79. Thus while there were differences for some schools overall the differences were not that substantial.

< ----- >
Insert Figure 5 about here
< ----- >

For the other pupil characteristics (ethnicity, FSM and gender) there was no significant school level variation in outcomes^{xiii}. Consider the FSM gap in progress for average age 11 score as shown in Table 7. Overall pupils on FSM made -.25 SD less progress than non-FSM pupils, and while this varies (although not significantly) across schools, there is still no school that eliminates the gap, the 95% CI range being from -.33 SD to -.16 SD. Indeed estimates of school effects on progress calculated separately for FSM and Non-FSM pupils^{xiv} correlate 0.98, as illustrated in Figure 6. Thus schools where pupils on FSM make the most progress are also the schools where pupils not on FSM make the most progress. Correlations for FSM and non-FSM pupils for the individual subjects were 0.96 for English and 0.99 for both maths and science.

< ----- >
Insert Figure 6 about here
< ----- >

The pattern holds for ethnic gaps. Ethnicity was reduced to the three main groups (White British, Black African and Black Caribbean) with a fourth category for all other ethnic groups combined. All four ethnic coefficients were allowed to vary simultaneously at the school level. In terms of the Black Caribbean-White British gap, overall Black Caribbean pupils made - 0.28 SD less progress than White British pupils, and while this varies (although not significantly) across schools, there is still no school that eliminates this gap, the 95% CI range

being from -0.34 SD to -0.21 SD (see Table 7). The correlation between school residual for White British pupils and for Black Caribbean pupils was 0.91, indicating that in schools where White British pupils made the greatest progress Black Caribbean pupils also made the greatest progress. This is illustrated in Figure 7. The correlation between school CVA residuals for White British and Black African pupils was 0.99, as was the correlation between the residuals for Black African and Black Caribbean pupils.

< ----- >
 Insert Figure 7 about here
 < ----- >

Finally the gender gap also did not vary significantly across schools with 95% confidence intervals across schools of 0.15 to 0.19 SD for English, -0.32 to -0.10 SD for maths and -0.20 to -0.02 SD for science. The correlations between schools CVA residuals for boys and for girls were again high in fact in excess of 0.99 for all subjects, indicating the schools where boys made the most progress were also the schools where girls made the most progress.

For completeness the analyses were repeated with fixed effects for year but without the hierarchical year level to model the data as a simple three year aggregation of pupils within schools. This did not change the conclusions drawn above. In summary, there is some evidence of differential school effectiveness in relation to prior attainment at age 7, with some schools having better CVA for pupils with low prior attainment than those with high prior attainment and other schools vice versa, although the correlation of 0.79 between CVA for low and high prior attainment suggests overall these differential results are not large. However there is no evidence of differential school effectiveness in relation to FSM, ethnicity or gender. It might be argued that the results reflect Bayesian shrinkage towards the mean in multi-level models, but replication of the results in OLS models revealed wider standard deviations but typical correlations between OLS and MLM estimates of around 0.97.

(iv) Differential effectiveness across models

Finally we ask how well the estimates of school effects correlate across the different models (Raw, VA and CVA) and assess this for different subjects and across years. The results are presented in Table 8 which gives for each subject both the lowest and the highest correlations found in any single year, as well as the average correlation across all three years. The correlations are surprisingly high. Even for the combination where we might expect the lowest correlation, between the raw and CVA models, the lowest correlation is still $r=0.75$ (for English in 2005). For maths and science the lowest correlations are $r=0.82$ and $r=0.84$ respectively. These correlations are not perfect and there are considerable changes for some schools. For example Figure 8 plots the schools' raw and CVA scores for English in 2005. We can see that of two schools with very low raw scores (both -0.60 SD) one also has a very low CVA score (-0.86 SD) but in the other pupils are making the progress expected given their prior attainment and background. What is indisputable though is that the correlations between schools' VA and CVA scores are near perfect, the absolute lowest being $r=0.94$ between schools VA and CVA for mathematics in 2004 (plotted in Figure 9). While CVA models are clearly superior in explaining significant additional amounts of the variance in pupil progress relative to simple VA models (see Table 4), they make very little difference to the rank order of schools.

< ----- >
 Insert Table 8, Figure 8 and Figure 9 about here
 < ----- >

DISCUSSION

Equity gaps in achievement and progress

While some authors have concluded that simple additive models are sufficient to reflect achievement gaps related to ethnicity, gender and poverty (e.g. Connolly, 2006; Rothson, 2007) the current results support research identifying significant interactions between these factors (e.g. Strand, 1999, 2010, 2014a,b; Kingdon & Cassen, 2010; Dekkers et al. 2000). The results show that White British pupils on FSM make less progress than their peers not on FSM, with a widening of the FSM gap over time. Poorer progress by pupils entitled to FSM compared to those not entitled has also been observed between age 4-7 (Strand, 1999) and is a notable feature of national data in England at all stages 7-11, 11-14 and 14-16 (DCSF, 2009; Strand, 2010, 2011). While the current data is drawn from a specific London LA, similar results have been reported in other London LAs (Strand 1999, 2014b) and in national data for all 11 year olds and 16 year olds in England (Strand, 2010, 2014a). The results also show that Black pupils, though particularly Black Caribbean pupils, make less progress than their White British peers among those not entitled to FSM. Of course entitlement to FSM is a blunt instrument offering no differentiation within the two-thirds of pupils not entitled to FSM, and it is possible that the Black Caribbean-White British gap among non-FSM pupils reflects unmeasured differences in other socio-economic variables such as family income, parental education, educational resources in the home or levels of social and cultural capital (e.g. Mandara, Varner, Greene & Richman, 2009; Kiernan & Mensah, 2011; Strand & Winston, 2008)^{xv}. Cultural factors may also be important in understanding differences in achievement between Black Caribbean and Black African pupils, with the latter group a more recent wave of migrants to England who, despite high levels of poverty, report the most positive attitudes to school, the highest levels of motivation and effort and have parents with high levels of educational qualifications and very high educational aspirations for their children (e.g. Strand, 2011; Strand & Winston, 2008). However it is still troubling that the FSM and Black Caribbean gaps appears to widen further between age 7-11. Such findings have been termed 'Matthew effects' (Stanovich, 1986) paraphrased as "to those that already hath shall be given...". Of course just because these gaps widen during a period when pupils are attending schools does not mean that schools themselves are the cause of widening gaps, we explore this further below.

School effects on equity gaps

School effects in this study were relatively large with a 0.50 SD difference between the less and more effective schools in CVA models adjusting for prior achievement and pupil and school background (although school effects were smaller in English). The school level accounted for 7% (English), 12% (mathematics) and 15% (science) of the residual variance in pupil progress, typical of the 5%-18% range reported in a recent review (Sammons, 2007). The factors that lie behind this variation between schools are not captured by the explanatory variables we have access to here, but may reflect factors such as leadership and management, school ethos, or quality of teaching (e.g., Harris, Chapman, Muijs, Russ & Stoll, 2006). Some authors argue that such differences between schools are central to both SES and ethnic achievement gaps (e.g., Fryer & Levitt, 2004; Wilson, Burgess & Briggs, 2011; Kingdon & Cassen, 2010). We do not argue that between school factors are irrelevant to equity gaps. For example the current results indicate there is negative association between achievement and the proportion of disadvantaged pupils in a school, with lower average

progress in schools with a high proportion of pupils entitled to FSM (ES= -0.18 SD) even after accounting for individual pupil's FSM status. Thus there are additional challenges in schools where the percentage of pupils on FSM is high. Other studies have reported similar outcomes (e.g. Dumay & Dupriez, 2008; Van Ewick & Slegers, 2010; Strand, 2010). The mechanisms through which SES composition effects are mediated are debated, but recent research in secondary schools from the US suggests that peer influences are a major mediating factor. Pupils' reports of whether a friend had dropped out of high school or whether the pupils' closest friend at school desired the pupil to attend college, could explain a substantial proportion of the SES composition effect, and were more predictive than school practices such as academic press, mean hours of homework or pupils' judgments of teacher quality (Palardy, 2013).

However whatever the factors that lie behind the SES compositional effect, or indeed the large amount of unexplained variation we observe between schools, the most important finding here is the lack of evidence of differential school effects on the progress of pupils in relation to pupil background variables such as FSM, ethnicity or gender. No schools reliably reversed the usual 'within school' pattern of FSM pupils making less progress than non-FSM pupils, or Black Caribbean pupils making less progress than White British pupils. Indeed primary schools that added the most value for White British pupils, girls or those not on FSM also added the most value for Black Caribbean pupils, boys and pupils on FSM. Strand (2010) noted the same results in an analysis of progress age 7-11 using national data from 534,000 pupils attending over 14,000 primary schools. Strand noted that the absence of statistically significant school variation in that study might be a reflection of the relatively small number of pupils within schools in any one year (average $n=37$), but the replication here with data collected over three years and an average within-school sample of $n=120$ suggests this is not the case. This also supports the earlier conclusion by Strand (1999) who found no differential school effects on pupil progress age 4-7 for a sample of 5,000 pupils from 55 primary schools with data collected over three successive cohorts. Overall we conclude that schools are unlikely to be the major locus of equity gaps in pupils' educational progress. While the FSM and the White British-Black Caribbean gaps widen between age 7 and age 11, these gaps do not appear to be significantly greater in some schools than in others, making it difficult to sustain an argument that it is due to individual 'school quality'.

We can in fact demonstrate the flaw in the 'school quality' argument with some quite simple data. In 2013 the England schools Inspectorate, OFSTED, published a summary of the achievement of FSM and non-FSM pupils in secondary schools in England in 2012 broken down by the OFSTED judgment of the schools' overall effectiveness ('Outstanding', 'Good', 'Requires Improvement' or 'Inadequate'). In schools judged by OFSTED as 'Outstanding' 50% of FSM pupils achieved the benchmark of 5 or more GCSE A*-C grades or equivalent including English and Maths (5EM) compared to just 25% of FSM pupils in schools judged by OFSTED to be 'Inadequate'. On the face of it this is good news: good schools can and do make a difference to the achievement of disadvantaged pupils. However while the outstanding schools 'raise the bar' they do not 'close the gap'. For example the odds of Non-FSM pupils achieving the 5EM benchmark are 2.7 times higher than the odds for FSM pupils within 'Inadequate' schools, but this is no better in the other three ratings; Indeed within schools rated by OFSTED as 'Outstanding' the odds for Non-FSM pupils achieving the 5EM benchmark are 3.0 times higher than FSM pupils. Bluntly speaking even if we improved all 'Inadequate' schools to the level of those judged 'Outstanding' we would still have a FSM gap and of much the same size as it is today. Figure 10 presents the data.^{xvi}

< ----->
 Insert Figure 10 about here
 < ----->

So if the gaps are not the results of variation in individual school policies, practice or effectiveness, what is driving the gaps? At the simplest level, large SES and ethnic gaps are observed at age 3 or 4 before children even start school (e.g. Strand, 1999b; Sylva et al, 2004; Bradshaw, 2011). Thus the major influences largely lie outside school themselves. Children who grow up in poverty may do less well because their parents are more stressed, less able to afford educational activities and resources, less well-placed to help them with their school work, more likely to experience alcohol or drug dependency, more likely live in unsafe and high crime areas and so on. These early disadvantages can accumulate like Matthew effects as new learning builds on old learning in a cumulative manner such that "the more progress made at any point in time the better the ability to profit from subsequent experiences" (Anastasi, 1976, p328). It may be that school processes are important in sustaining achievement gaps, but the focus needs to be on widespread *within-school* processes. For example recent research suggests that setting (or tracking as it is termed in the US) is quite widespread within primary schools in England, with a recent nationally representative survey indicating that over one-third of children in Y2 were set in literacy, mathematics or both, and that low SES children were disproportionately represented in bottom sets (Hallam & Parsons, 2013). Setting has been hypothesised to increase within school gaps for disadvantaged and ethnic minority pupils (Gillborn & Youdell, 2000, William & Bartholomew, 2004; Oakes, 2005; Schofield, 2010). Another possible candidate is the distribution of teachers across classrooms *within* schools, with research from the US indicating that Black and disadvantaged pupils were more likely to be matched to novice teachers while more highly qualified teachers were matched to White and more advantaged pupils (Clotfelter, Ladd & Vigor, 2005, 2006). They suggest this reflects both pressure from more affluent parents to secure better resources for their children and the greater power of more experienced teachers to lobby for more easy-to-educate pupils. Other possibilities, such as widespread low expectations on the part of teachers for White working class and Black Caribbean pupils, cannot be discounted (e.g. Reay, 2006; Gillborn, 2008; Strand 2012).

Whatever the reasons, and there are likely to be many rather than one, our results have significant implications for policy, which needs to focus on within-school gaps rather than on school structures. We know that some schools in disadvantaged areas 'succeed against the odds' (e.g. National Commission on Education, 1996; DCSF Extra Mile project, 2008). On this basis Her Majesty's Chief Inspector (HMCI) in a speech to the National College of School Leaders argued "*If some schools and communities can successfully narrow the gap, why can't others do so too?*" (Wilshaw, 2012). In the same vein Alan Milburn, Chair of the Social Mobility and Child Poverty Commission, was recently quoted as follows "*Some schools are proving that deprivation needn't be destiny. They have cracked the code on how to improve social mobility by helping disadvantaged children to excel in education. If some schools can do it, there is no excuse for others not to*" (Burn, 2014). However such schools are by definition exceptions. As Mortimore and Whitty (1997) observe "*whilst some schools can succeed against the odds, the possibility of them all doing so, year in and year out, still appears remote given that the long-term patterning of educational inequality has been strikingly consistent throughout the history of public education in most countries*" (p9). Similarly in the US Rothstein (2004) has cogently argued that, exemplary though such schools might be, they do not represent scalable models for America's poor and minority children (for an England

perspective see Barker, 2010). As we have seen here, while the most effective schools may raise the achievement of disadvantaged pupils, non-disadvantaged pupils may benefit just as much, so that these schools do not eliminate the achievement gap. Purely school focussed interventions, such as closing schools falling below 'floor targets' and reopening them as academies, or opening new 'Free' or 'Charter' schools, may have limited success unless they also tackle poverty, low aspirations, the home learning environment and other factors outside school. This indicates the need for positive discrimination and greater targeting of human and material resources. The introduction of the Pupil Premium in April 2011 designed to target extra funding at deprived pupils (those entitled to FSM or Looked After by the Local Authority), and the focus on the FSM gap *within* all schools, may have the power to deliver greater change, though clear evidence has not yet emerged (National Audit Office, 2015).

Differential effectiveness across subjects, over time and between models

The above concern to understand differential schools effects in relation to equity gaps was the main driver for this analysis, but the results also allow us to draw conclusions about differential school effects on other dimensions such as curriculum subjects and over time. Primary schools' CVA results were moderately stable across subjects, with correlations ranging from 0.65 between English and maths to 0.83 between maths and science. While these are moderate to high correlations, indicating an overall summary score has reasonable validity, there is enough variation to indicate that some schools may do very well say in English but poorly in mathematics or vice versa (e.g. see Figure 3). Correlations in CVA scores over time were moderate for maths and science, around 0.60 for scores two years apart, but were exceptionally low for English at 0.23. This indicates the importance of calculating three year rolling averages for each school which should substantially increase the stability of school estimates, and indeed the DFE instigated this for the 2013 performance tables. Rolling averages are also important if reporting results by pupil characteristics such as prior attainment or entitlement to FSM, since the DFE will not publish scores where groups contain five or fewer pupils and this is not uncommon in primary schools. The inclusion in 2013 performance tables of value added measures separately for pupils with low, average and high prior attainment is also supported by the current results, since this was the only dimension where we found (very small) differential school effects. While we did not find differential school effects by pupils' FSM status we do support the Governments move to publish schools' results separately for pupils entitled to FSM and those not entitled to FSM. Although these scores are extremely highly correlated across schools it does serve to focus the attention of all schools' on their FSM gap, not just those schools below floor targets^{xvii}.

It is notable that there are relatively high correlations between raw and CVA school residuals which ranged from 0.78 for English to 0.82 for maths and 0.86 for science. The results reported here are not unusual and similar estimates have been reported before (e.g. Schagen & Styles, 2006). Such relatively high correlations have led some authors to criticize CVA measures as no more than "the transformed *equivalent* of raw scores" (Gorard, 2006, p241, original italics). We would though expect a correlation between any CVA measure and raw scores since schools that facilitate better pupil progress would in general have higher raw score outcomes. Also, as Schagen & Styles (2006) observe, because two variables are correlated (like height and weight) does not mean they are measuring the same thing. Finally, the correlation between VA and CVA measures reported here is high, over 0.94, but CVA explains more variation and has consistently lower correlations with raw results than VA, indicating accounting for pupil background and school composition is important. It is regrettable that CVA measures were removed from performance tables in 2011, particularly

since the simple threshold VA measure that replaced them (the percentage of pupils making 2+ levels of progress) is highly problematic and has many limitations compared even just to more sophisticated VA measures based on average scores (see Leckie, 2015).

Conclusions

A strong focus on schools as the key driver in addressing achievement gaps has characterised Government policy over the last 25 years, and explanations that identify 'low quality' schools as the cause of poverty or ethnic achievement gaps continue to be espoused. This explanation is seductive since it suggests the problem resides in a minority of failing schools which, if these schools can be fixed, will ameliorate the problem. However equity gaps are large and substantial before children start school and do not appear to be significantly greater in some schools than in others, suggesting the gaps are a systemic issue rather than the result of a small number of 'failing' schools. This is not to say that schools cannot play a part in reducing achievement gaps, but to position them as the prime movers is at best a misreading of the evidence and at worse a distraction from the real issues.

REFERENCES

- Anastasi, A. (1976). *Psychological testing (4th Ed.)*. New York: Macmillan.
- Bali, V., & Alvarez, R. (2004). The Race Gap in Pupil Achievement Scores: Longitudinal Evidence from a Racially Diverse School District. *Policy Studies Journal*, 32, (3), 393-415.
- Barker, B. (2010). *The pendulum swings: transforming school reform*. Stoke on Trent: Trentham Books.
- Bernstein, B. (1970). Education cannot compensate for society. *New Society*, 387, 344-347.
- Boudon, R. (1974). *Education, opportunity and social inequality*. New York: Wiley.
- Bradshaw, P. (2011). *Growing up in Scotland: Changes in child cognitive ability in the pre-school years*. Edinburgh: Scottish Government.
- Brandsma, H. P., & Knuver, J. W. (1989). Effects of school and classroom characteristics on pupil progress in language and arithmetic. *International Journal of Educational Research*, 13(7), 777-788.
- Burns, J. (2014). *Schools 'must do more' to help disadvantaged pupils*. BBC News, 6 October 2014. <http://www.bbc.co.uk/news/education-29478935>
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. (2005). Who teaches whom? Race and the distribution of novice teachers. *Economics of Education Review*, 24(4), 377-392.
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-Pupil Matching and the Assessment of Teacher Effectiveness. *Journal of Human Resources*, XLI(4), 778-820.
- Connolly, P. (2006). The effects of social class and ethnicity on gender differences in GCSE Attainment: A secondary analysis of the youth cohort study of England and Wales 1997-2001. *BERJ*, 32, (1), 3-21. *British Educational Research Journal*, 32(1), 3-21.
- Day, C., Stobart, G., Sammons, P., Kingon, A., Gu, Q., & Mujtaba, T. (2006). *Variations in teachers' work, lives and effectiveness*. Nottingham: Department for Education and Skills.
- DCSF (2007). *Gender and education: the evidence on pupils in England*. London: Department for Children Schools and Families.
- DCSF (2008). *The extra mile: How schools succeed in raising aspirations in deprived communities*. Ref 00447-2008BKT-EN. London: Department for Schools, Children and Families.
- DCSF (2009). *Deprivation and education: The evidence on pupils in England, Foundation Stage to Key Stage 4*. London: Department for Children Schools and Families.

- De Maeyer, S., van den Bergh, H., Rymenans, R., Van Petegem, P., & Rijlaarsdam, G. (2010). Effectiveness criteria in school effectiveness studies: Further research on the choice for a multivariate model. *Educational Research Review*, 5(1), 81-96.
- Dearden, L., Micklewright, J., & Vignoles, A. (2011). The Effectiveness of English Secondary Schools for Pupils of Different Ability Levels. *Fiscal Studies*, 32(2), 225-244.
- Dekkers, H., Bosker, R., & Driessen, G. (2000). Complex inequalities of educational opportunities: A large scale longitudinal study on the relation between gender, social class, ethnicity and school success. *Educational Research and Evaluation*, 6(1), 59-82.
- DES (1985). Education for all. The report of the committee of enquiry into the education of children from ethnic minority groups (Swann Report). London: Her Majesty's Stationary Office (HMSO).
- DFE (2011). *National Curriculum assessments at Key Stage 2 in England, 2011 (Revised)*. Available from the www at <http://www.education.gov.uk/rsgateway/DB/SFR/s001047/index.shtml> [Accessed 1 September 2015].
- DFE (2014a). *DFE Consolidated Annual Report and Accounts 2012-13*. London: Department for Education.
- DFE (2014b). *Reforming Assessment and Accountability for Primary Schools: Government response to consultation on primary school assessment and accountability*. London: Department for Education.
- DfES (2006). *Ethnicity and education*. London: DfES. Available from the world wide web at: <http://webarchive.nationalarchives.gov.uk/20130401151715/http://www.education.gov.uk/publications/eOrderingDownload/DFES-0208-2006.pdf> [accessed 1 September 2015].
- Douglas, J. (1964). *The home and the school: A study of ability and attainment in the primary school*. London: MacGibbon & Kee.
- Dumay, X., & Dupriez, V. (2008). Does the school composition effect matter? Evidence from Belgian data. *British Journal of Educational Studies*, 56(4), 440-477.
- Dumay, X., Coe, R., & Anumendem, D. N. (2014). Stability over time of different methods of estimating school performance. *School Effectiveness and School Improvement*, 25(1), 64-82.
- Fryer, R., & Levitt, S. (2004). Understanding the Black-White test score gap in the first two years of school. *Review of Economics and Statistics*, 86(2), 447-464.
- Fryer, R., & Levitt, S. (2006). The Black-White Test Score Gap Through Third Grade. *American Law and Economics Review*, 8, (2), 249-281.
- Gillborn, D. (2008). Coincidence or conspiracy? Whiteness, policy and the persistence of the Black/White achievement gap. *Educational Review*, 60(3), 229 - 248.
- Gillborn, D., & Youdell, D. (2000). *Rationing education: Policy, practice, reform and equity*. Buckingham: Open University Press.
- Goldstein, Harvey, & Noden, Philip. (2004). A response to Gorard on social segregation. *Oxford Review of Education*, 30(3), 441-442.
- Gorard, S. (2006). Value-added is of little value. *Journal of Education Policy*, 21(2), 235-243.
- Gray, J., Goldstein, H., & Thomas, S. (2001). Predicting the future: The role of past performance in determining trends in institutional effectiveness at A level. *British Educational Research Journal*, 27(4), 391-405.
- Hallam, S., & Parsons, S. (2013). The incidence and make up of ability grouped sets in the UK primary school. *Research Papers in Education*, 28(4), 393-420.
- Harris, A., Chapman, C., Muijs, D., Russ, J., & Stoll, L. (2006). Improving schools in challenging contexts: Exploring the possible. *School Effectiveness and School Improvement*, 17(4), 409-424.

- He, Q., Hayes, M., & Wiliam, D. (2013). Classification accuracy in Key Stage 2 National Curriculum tests in England. *Research Papers in Education*, 28(1), 22-42.
- Hill, P., & Rowe, K. (1996). Multilevel Modelling in School Effectiveness Research. *School Effectiveness and School Improvement*, 7(1), 1-34.
- House of Commons Education Committee (2015). *Academies and free schools: Fourth report of Session 2014-15 (HC 258)*. London: The Stationary Office.
www.parliament.uk/education-committee.
- Huang, H., & Sebastian, J. (2015). The role of schools in bridging within-school achievement gaps based on socioeconomic status: A cross-national comparative study. *Compare*, 45(4), 501-525.
- Jaccard, J., & Turrissi, R. (2003). *Interaction effects in multiple regression* (second edition). Thousand Oaks: Sage.
- Kiernan, K. E., & Mensah, F. K. (2011). Poverty, family resources and children's early educational attainment: the mediating role of parenting. *British Educational Research Journal*, 37(2), 317-336.
- Kingdon, G., & Cassen, R. (2010). Ethnicity and low achievement in English schools. *British Educational Research Journal*, 36(3), 403-431.
- Kyriakides, L. (2004). Differential School Effectiveness in Relation to Sex and Social Class: Some Implications for Policy Evaluation. *Educational Research and Evaluation*, 10, (2), 141 - 161.
- Leckie, G. (2013). *Three-Level Multilevel Models - Concepts*. LEMMA VLE Module 11, 1-47.
<http://www.bristol.ac.uk/cmm/learning/course.html>
- Leckie, G. (2015). *Monitoring school performance: A statistical critique of 'expected progress' with comparison to multilevel 'value-added' models*. University of Bristol.
- Leckie, G., & Goldstein, H. (2011). Understanding Uncertainty in School League Tables. *Fiscal Studies*, 32(2), 207-224.
- Luyten, H. (1998). School effectiveness and pupil achievement, consistent across subjects? Evidence from Dutch elementary and secondary education. *Educational Research and Evaluation*, 4(4), 281-306.
- Ma, X. (2000). Socioeconomic gaps in academic achievement within schools: Are they consistent across subjects? *Educational Research and Evaluation*, 6(4), 337-355.
- Mandara, J., Varner, F., Greene, N., & Richman, S. (2009). Intergenerational Family Predictors of the Black-White Achievement Gap. *Journal of Educational Psychology*, 101(4), 867-878.
- Mangan, J., Pugh, G. and Gray, J. (2005). Changes in examination performance in English secondary schools over the course of a decade, *School Effectiveness and School Improvement*, 16(1), 29-50.
- McIntosh, S., & Vignoles, A. (2000). *Measuring and Assessing the Impact of Basic Skills on Labour Market Outcomes*. London: Centre for the Economics of Education.
- Melhuish, E., Romaniuk, H., Sammons, P., Sylva, K., Siraj-Blatchford, I., & Taggart, B. (2006). *The effectiveness of primary schools in England in Key Stage 2 for 2002, 2003 and 2004*. London: Institute of Education.
- Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters: The junior years*. Wells Somerset, UK: Open Books.
- Mortimore, P., & Whitty, G. (1997). *Can school improvement overcome the effects of disadvantage?* London: Institute of Education.
- National Audit Office (2015). *Funding for disadvantaged pupils: Report by the Comptroller and Auditor General*. London: Stationary Office.

- National Commission on Education (1996). *Success against the odds: Effective schools in disadvantaged areas*. London: Routledge.
- Nuttall, D., Goldstein, H., Prosser, R., & Rasbash, J. (1989). Differential school effectiveness. *International Journal of Educational Research*, 13, 769-776.
- Oakes, J. (2005). *Keeping Track: How Schools Structure Inequality (2nd Edition)*. New Haven: Yale University Press.
- OFSTED. (2013). *Unseen children: Access and achievement 20 years on*. London: Office for Standards in Education.
- Palardy, G. J. (2008). Differential school effects among low, middle, and high social class composition schools: a multiple group, multilevel latent growth curve analysis. *School Effectiveness and School Improvement*, 19(1), 21 - 49.
- Palardy, G. J. (2013). High School Socioeconomic Segregation and Pupil Attainment. *American Educational Research Journal*, 50(4), 714-754.
- Phillips, M. Crouse, J. & Ralph (1998). *Does the Black White test score gap widen after children enter school?* In Jencks, C. & Philips, M. (1998). *The Black-White test score gap*. Washington: Brookings Institution Press.
- Plewis, Ian, & Fielding, Antony. (2003). What is Multi-Level Modelling for? A Critical Response to Gorard (2003). *British Journal of Educational Studies*, 51(4), 408-419
- Reay, D. (2006). The zombie stalking English schools: Social class and educational inequality. *British Journal of Educational Studies*, 54(3), 288-307.
- Rothson, C. (2007). Can achievement differentials be explained by social class alone? An examination of minority ethnic educational performance in England and Wales at the end of compulsory schooling. *Ethnicities*, 7(3), 306-322.
- Rothstein, R. (2004). *Class and Schools: Using Social, Economic, and Educational Reform to Close the Achievement Gap*: Economic Policy Institute, Washington.
- Rutter, M., Maughan, B., Mortimore, P., Ouston, J., & Smith, A. (1979). *Fifteen thousand hours: Secondary schools and their effects on children*. Cambridge, MA: Harvard University Press.
- Sammons, P. (2007). *School effectiveness and equity: Making connections*. Reading: CfBT Trust. Available from the WWW at <http://www.cfbt.com/evidenceforeducation/pdf/Full%20Literature%20Review.pdf> [Accessed 11 July 2012].
- Sammons, P., Nuttall, D., & Cuttance, P. (1993). Differential School Effectiveness: Results from a Reanalysis of the Inner London Education Authority's Junior School Project Data. *British Educational Research Journal*, 19(4), 381-405.
- Schagen, I., & Styles, B. (2006). Value-added analysis is useful: A response to Gorard. *Research Intelligence*, 96, p9.
- Schofield, J. W. (2010). International evidence on ability grouping with curriculum differentiation and the achievement gap in secondary schools. *Teachers College Record*, 112(5), 1492-1528.
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417-453.
- Smith, D. J., & Tomlinson, S. (1989). *The school effect: A study of multi-racial comprehensives*. London: Policy Studies Institute.
- Stanovich, K. E. (1986). Matthew Effects in Reading: Some Consequences of Individual Differences in the Acquisition of Literacy. *Reading Research Quarterly* 21(4), 360-407.
- Strand, S. (1997). Pupil Progress during Key Stage 1: a value added analysis of school effects. *British Educational Research Journal*, 23(4), 471 - 487.

- Strand, S. (1999). Ethnic group, sex and economic disadvantage: Associations with pupils' educational progress from Baseline to the end of Key Stage 1. *British Educational Research Journal*, 25, (2), 179-202.
- Strand, S. (1999). Baseline assessment results at age 4: associations with pupil background factors. *Journal of Research in Reading*, 22(1), 14-26.
- Strand, S. (2010). Do some schools narrow the gap? Differential school effectiveness by ethnicity, gender, poverty and prior attainment. *School Effectiveness and School Improvement*, 21, (3), 289-314.
- Strand, S. (2011). The limits of social class in explaining ethnic gaps in educational attainment. *British Educational Research Journal*, 37, (2), 197-229.
- Strand, S. (2012). The White British–Black Caribbean achievement gap: Tests, tiers and teacher expectations. *British Educational Research Journal*, 38(1), 75-101.
- Strand, S. (2014a). Ethnicity, gender, social class and achievement gaps at age 16: intersectionality and 'getting it' for the white working class. *Research Papers in Education*, 29(2), 131-171.
- Strand, S. (2014b). School effects and ethnic, gender and socio-economic gaps in educational achievement at age 11. *Oxford Review of Education*, 40(2), 223-245.
- Strand, S., & Winston, J. (2008). Educational aspirations in inner city schools. *Educational Studies*, 34(4), 249-267.
- Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2004). *The Effective Provision of Pre-School Education (EPPE) Project: Final Report*. Nottingham: Department for Education and Skills.
- Thomas, S. (2001). Dimensions of Secondary School Effectiveness: Comparative Analyses Across Regions. *School Effectiveness and School Improvement*, 12(3), 285 - 322.
- Thomas, S., Sammons, P., Mortimore, P., & Smees, R. (1997a). Stability and consistency in secondary schools effects on pupils' GCSE outcomes over three years. *School Effectiveness and School Improvement*, 8(2), 169-197.
- Thomas, S., Sammons, P., Mortimore, P., & Smees, R. (1997b). Differential Secondary School Effectiveness: Comparing the Performance of Different Pupil Groups. *British Educational Research Journal*, 23(4), 451-469.
- Tymms, P. (2004). Effect sizes in multilevel models. In I. Schagen & K. Elliott (Eds.), *But what does it mean? The use of effect sizes in educational research*. (pp. 55-66). Slough: National Foundation for Educational Research.
- Van Ewijk, R., & Slegers, P. (2010). The effect of peer socioeconomic status on pupil achievement: A meta-analysis. *Educational Research Review*, 5(2), 134-150.
- Wiliam, D., & Bartholomew, H. (2004). It's not which school but which set you're in that matters: The influence of ability grouping practices on pupil progress in mathematics. *British Educational Research Journal*, 30(2), 279-293.
- Wilson, D., Burgess, S., & Briggs, A. (2011). The dynamics of school attainment of England's ethnic minorities. *Journal of Population Economics*, 24(2), 681-700.
- Wilshaw, M. (2012). HMCI speech to the National College of School leaders, 15 June 2012.

Table 1: Descriptive statistics for the sample

Variable	Value	N	%	Age 7 score (normalised)		Age 11 scores (normalised)					
				Mean	SD	English		maths		science	
						Mean	SD	Mean	SD	Mean	SD
Ethnic group	White Other Groups	399	6.6%	.02	(1.03)	.24	(.97)	.27	(.98)	.28	(.97)
	Portuguese	300	5.0%	-.67	(.84)	-.29	(.86)	-.15	(.93)	-.21	(.93)
	Mixed White & Caribbean	356	5.9%	.12	(.89)	.10	(.84)	.11	(.95)	.17	(.91)
	Mixed Other Heritage	333	5.5%	.29	(1.01)	.38	(.96)	.33	(1.01)	.40	(.97)
	Black African	1281	21.2%	-.01	(.91)	.12	(.89)	.10	(.91)	.01	(.91)
	Black Caribbean	1361	22.5%	-.15	(.91)	-.14	(.88)	-.22	(.88)	-.16	(.88)
	Black Other Groups	256	4.2%	-.12	(.94)	.00	(.95)	-.08	(.93)	-.06	(.86)
	Bangladeshi/Pakistani	170	2.8%	-.11	(1.02)	.02	(.84)	.04	(.94)	-.10	(.80)
	Other Asian	190	3.1%	.26	(.86)	.27	(.83)	.57	(1.01)	.33	(.97)
	Any Other group	200	3.3%	.01	(1.05)	.15	(.93)	.22	(.96)	.18	(.95)
	White-British	1199	19.8%	.20	(1.03)	.28	(1.07)	.27	(1.00)	.34	(1.04)
Free School Meals	Not entitled FSM	3723	61.6%	.17	(.95)	.27	(.95)	.25	(.95)	.26	(.96)
	Entitled FSM	2322	38.4%	-.29	(.94)	-.20	(.88)	-.19	(.92)	-.20	(.89)
gender	Boy	2955	48.9%	-.13	(.99)	-.10	(.92)	.08	(1.00)	.04	(.98)
	Girl	3090	51.1%	.11	(.94)	.28	(.94)	.09	(.93)	.13	(.94)
Age in months (deviated from mean=134)	-5	519	8.6%	-.22	(.96)	-.03	(.96)	-.02	(.95)	.03	(.95)
	-4	483	8.0%	-.20	(.91)	.00	(.91)	-.03	(.94)	.00	(.96)
	-3	497	8.2%	-.21	(.91)	-.06	(.91)	-.03	(.91)	-.03	(.90)
	-2	515	8.5%	-.08	(.97)	.05	(.93)	.04	(.93)	.05	(.94)
	-1	450	7.4%	-.13	(.96)	.01	(.95)	-.03	(.93)	-.05	(.96)
	0	538	8.9%	-.06	(.97)	.09	(.99)	.08	(1.01)	.10	(1.01)
	1	464	7.7%	.04	(.94)	.10	(.95)	.10	(.98)	.07	(.93)
	2	488	8.1%	.03	(1.00)	.06	(.96)	.08	(.96)	.08	(.98)
	3	502	8.3%	.05	(1.04)	.07	(.95)	.06	(1.03)	.08	(.99)
	4	489	8.1%	.16	(.95)	.21	(.96)	.17	(.94)	.20	(.95)
	5	540	8.9%	.24	(.96)	.24	(.96)	.25	(.99)	.20	(.97)
	6	560	9.3%	.29	(.95)	.32	(.91)	.28	(.93)	.25	(.93)
Year joined school	Y3 (Jan. on)	104	1.7%	-.20	(1.06)	-.13	(.98)	.02	(.99)	-.05	(.96)
	Y4	278	4.6%	-.21	(.99)	-.13	(.97)	-.20	(1.00)	-.08	(.99)
	Y5	296	4.9%	-.19	(.92)	-.05	(.91)	-.09	(.99)	-.11	(.95)
	Y6	138	2.3%	-.34	(.91)	-.47	(.98)	-.38	(1.02)	-.47	(1.00)
	Stable	5229	86.5%	.03	(.97)	.13	(.94)	.12	(.95)	.12	(.95)
SEN	School Action	849	14.0%	-.58	(.77)	-.44	(.72)	-.45	(.74)	-.35	(.83)
	School Action Plus	769	12.7%	-.88	(.77)	-.79	(.75)	-.73	(.82)	-.60	(.87)
	Statemented	179	3.0%	-1.34	(.91)	-1.23	(.92)	-1.12	(.97)	-1.05	(1.04)
	No SEN identified	4248	70.3%	.33	(.84)	.41	(.82)	.38	(.85)	.34	(.87)
Stage of Fluency in English Language	Beginner	8	0.1%	-.95	(1.11)	-1.21	(1.10)	-.83	(1.06)	-1.32	(1.24)
	Considerable support	233	3.9%	-1.25	(.74)	-.95	(.72)	-.83	(.79)	-.83	(.83)
	Some support	910	15.1%	-.42	(.80)	-.19	(.73)	-.10	(.86)	-.18	(.83)
	Fully fluent	1221	20.2%	.37	(.83)	.45	(.82)	.46	(.87)	.34	(.86)
	English only	3673	60.8%	.06	(.97)	.11	(.98)	.06	(.97)	.13	(.98)
Year	2004	1992	33.0%	-.02	(.99)	.10	(.95)	.09	(.96)	.09	(.96)
	2005	2003	33.1%	.00	(.96)	.09	(.95)	.08	(.97)	.08	(.96)
	2006	2050	33.9%	.01	(.98)	.09	(.95)	.08	(.97)	.09	(.96)
Grand total		6045	100%	.00	(.98)	.09	(.95)	.08	(.96)	.09	(.96)

Note: SEN= Special Educational Needs.

Table 2: Multi-level regression analyses for age 11 scores in English, maths, science and average age 11 score

Variable	Value	English		Maths		Science		Average score	
		Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Intercept		0.18	0.05	0.31	0.05	0.35	0.06	0.29	0.05
Year	2005 vs. 2004	0.00	0.05	0.02	0.04	-0.01	0.05	0.01	0.04
	2006 vs. 2004	-0.01	0.05	0.02	0.04	0.00	0.05	0.00	0.04
Age	Age in months (normalised)	0.00	0.00	-0.01 *	0.00	-0.01 *	0.00	-0.01 *	0.00
FSM	Entitled FSM (Vs. Not Entitled)	-0.27 *	0.03	-0.17 *	0.04	-0.27 *	0.04	-0.25 *	0.03
Gender	Girl (vs. Boy)	0.17 *	0.01	-0.21 *	0.02	-0.11 *	0.02	-0.06 *	0.01
Mobility	Joined Y3	0.02	0.05	0.11	0.06	0.03	0.06	0.06	0.05
	Joined Y4	-0.07 *	0.03	-0.07	0.04	0.02	0.04	-0.06	0.03
	Joined Y5	-0.02	0.04	-0.03	0.04	0.01	0.04	-0.02	0.03
	Joined Y6	-0.17 *	0.06	-0.13 *	0.06	-0.08	0.07	-0.15 *	0.06
	Stable (same school)	-	-	-	-	-	-		
Stage of fluency in English	Beginner/Considerable support	-0.27 *	0.05	-0.18 *	0.05	-0.30 *	0.05	-0.26 *	0.04
	Some support	-0.09 *	0.03	0.02	0.03	-0.10 *	0.04	-0.05	0.03
	Fully fluent in English	0.04	0.03	0.10 *	0.03	0.01	0.03	0.06 *	0.03
	Monolingual English	-	-	-	-	-	-		
SEN	School action	-0.27 *	0.02	-0.25 *	0.02	-0.17 *	0.03	-0.26 *	0.02
	School action plus	-0.42 *	0.02	-0.37 *	0.03	-0.26 *	0.03	-0.38 *	0.02
	Statemented	-0.61 *	0.05	-0.52 *	0.05	-0.45 *	0.05	-0.58 *	0.04
	No SEN	-	-	-	-	-	-		
Ethnic group	White Other Groups	0.05	0.04	0.05	0.05	0.04	0.05	0.04	0.04
	Portuguese	0.01	0.05	0.06	0.06	0.02	0.06	0.03	0.05
	Mixed White & Caribbean	-0.10 *	0.04	-0.10 *	0.05	-0.14 *	0.05	-0.13 *	0.04
	Mixed Other Heritage	-0.02	0.04	-0.06	0.05	-0.06	0.05	-0.05	0.04
	Black African	-0.03	0.03	-0.11 *	0.04	-0.16 *	0.04	-0.11 *	0.03
	Black Caribbean	-0.20 *	0.03	-0.26 *	0.03	-0.29 *	0.03	-0.28 *	0.03
	Black Other Groups	-0.06	0.05	-0.18 *	0.05	-0.26 *	0.06	-0.18 *	0.05
	Bangladeshi/Pakistani	-0.05	0.06	-0.14 *	0.07	-0.29 *	0.07	-0.17 *	0.06
	Other Asian	-0.01	0.06	0.16 *	0.06	0.00	0.07	0.06	0.06
	Any Other group	-0.06	0.05	0.04	0.06	0.02	0.06	-0.01	0.05
	White British	-	-	-	-	-	-		
Ethnic * FSM	White Other Groups * FSM	0.24 *	0.06	0.19 *	0.07	0.24 *	0.08	0.24 *	0.06
	Portuguese * FSM	0.21 *	0.07	0.06	0.08	0.18 *	0.09	0.15 *	0.07
	Mixed White & Caribbean * FSM	0.05	0.07	0.06	0.07	0.15	0.08	0.09	0.06
	Mixed Other Heritage * FSM	0.21 *	0.07	0.07	0.08	0.22 *	0.08	0.16 *	0.07
	Black African * FSM	0.23 *	0.05	0.13 *	0.05	0.19 *	0.05	0.20 *	0.04
	Black Caribbean * FSM	0.23 *	0.05	0.13 *	0.05	0.23 *	0.05	0.21 *	0.04
	Black Other Groups * FSM	0.16 *	0.08	0.18 *	0.08	0.38 *	0.09	0.25 *	0.07
	Bangladeshi/Pakistani * FSM	0.29 *	0.09	0.25 *	0.10	0.39 *	0.10	0.33 *	0.09
	Other Asian * FSM	0.15	0.09	0.23 *	0.09	0.18	0.10	0.19 *	0.08
	Any Other group * FSM	0.26 *	0.09	0.24 *	0.09	0.23 *	0.10	0.26 *	0.08
Prior attainment	age 7 score	0.57 *	0.01	0.61 *	0.01	0.54 *	0.01	0.63 *	0.01
	Age 7 score squared	0.06 *	0.01	0.06 *	0.01	0.04 *	0.01	0.07 *	0.01
School Composition	% FSM	0.00 *	0.00	-0.01 *	0.00	0.00 *	0.00	-0.01 *	0.00
	School mean age 7 score	-0.13	0.07	-0.27 *	0.07	-0.13	0.08	-0.21 *	0.07

*Note: Age 11 English, maths and science results tested in a multivariate model. Average age 11 score from a univariate model. FSM= Free School Meals; SEN= Special Educational Needs. *= $p < .05$.*

Table 3: Descriptive statistics for school-cohort composition variables

Composition variable	N	Mean	SD	5th percentile	95th percentile
Mean Y6 roll	171	40.2	15.2	23.0	67.0
% FSM	171	37.9	15.1	13.6	64.8
% joined Y3 or later	171	21.0	13.7	5.5	42.4
% not fluent in English	171	22.0	15.8	0.0	49.5
% girls	171	50.9	8.8	35.0	64.8
% White British	171	18.2	11.5	2.7	41.4
% Black African	171	23.1	13.5	4.7	50.0
% Black Caribbean	171	23.1	11.4	4.0	40.8
% SEN	171	29.9	11.6	10.3	50.0
Mean age	171	0.55	0.61	-0.52	1.49
Average age 7 score	171	-0.01	0.38	-0.67	0.55

Notes: Mean and Standard Deviation (SD) calculated from 171 observations.

Table 4: random part of the multivariate multi-level model

Random Part	Null Model				Value Added (VA)				Contextual Value Added (CVA)			
	Est.	SE	VPC	VR	Est.	SE	VPC	VR	Est.	SE	VPC	VR
School												
English	0.102	0.024	11.3%	-	0.037	0.012	8.9%	63.7%	0.023	0.009	6.5%	77.5%
Maths	0.113	0.024	12.1%	-	0.065	0.015	14.0%	42.5%	0.049	0.012	12.2%	56.6%
science	0.157	0.033	17.0%	-	0.091	0.021	17.1%	42.0%	0.074	0.018	15.2%	52.9%
Year												
English	0.051	0.010	5.6%	-	0.062	0.010	14.9%	-21.6%	0.060	0.009	16.9%	-17.6%
Maths	0.019	0.006	2.0%	-	0.027	0.005	5.8%	-42.1%	0.025	0.005	6.2%	-31.6%
science	0.035	0.008	3.8%	-	0.045	0.008	8.5%	-28.6%	0.044	0.007	9.1%	-25.7%
Pupil												
English	0.753	0.014	83.1%	-	0.317	0.006	76.2%	57.9%	0.273	0.005	76.7%	63.7%
Maths	0.799	0.015	85.8%	-	0.371	0.007	80.1%	53.6%	0.327	0.006	81.5%	59.1%
science	0.731	0.013	79.2%	-	0.395	0.007	74.4%	46.0%	0.368	0.007	75.7%	49.7%
Total												
English	0.906	-	-	-	0.416	-	-	54.1%	0.356	-	-	60.7%
Maths	0.931	-	-	-	0.463	-	-	50.3%	0.401	-	-	56.9%
science	0.923	-	-	-	0.531	-	-	42.5%	0.486	-	-	47.3%
Cohort estimates												
			CV	r			CV	r			CV	r
English			33.3%	0.67			62.6%	0.37			72.3%	0.28
Maths			14.4%	0.86			29.3%	0.71			33.8%	0.66
science			18.2%	0.82			33.1%	0.67			37.3%	0.63
-2*loglikelihood:												
Schools	35398				29284				27542			
Cohorts	57				57				57			
Students	171				171				171			
Response	6045				6045				6045			
	18092				18092				18092			

Notes: Model fitted with just an intercept term random at each level. Est.= estimated coefficient of intercept term; SE= standard error; VPC= Variance Partition Coefficient; VR= Variance reduction relative to the null model; CV= within-school cohort variance as a percentage of the total between school variance; r= average correlation between cohorts within schools. Covariance terms omitted for clarity of presentation.

Table 5: Correlations between school residuals for different subjects

Model	Subject	2004	2005	2006	Avg.
Null model	English vs. maths	0.81	0.80	0.80	0.80
	English vs. science	0.82	0.83	0.77	0.81
	Maths vs. science	0.88	0.91	0.89	0.89
VA model	English vs. maths	0.70	0.72	0.69	0.70
	English vs. science	0.74	0.77	0.61	0.71
	Maths vs. science	0.82	0.88	0.83	0.84
CVA model	English vs. maths	0.62	0.66	0.66	0.65
	English vs. science	0.71	0.74	0.56	0.67
	Maths vs. science	0.80	0.87	0.82	0.83

Table 6: Correlations between school residuals over time

Model	Subject	2004- 2005	2005- 2006	2004- 2006	Avg.
Null model	English	0.75	0.60	0.63	0.66
	Maths	0.89	0.89	0.80	0.86
	Science	0.86	0.81	0.79	0.82
VA model	English	0.47	0.38	0.31	0.38
	Maths	0.71	0.81	0.62	0.71
	Science	0.70	0.70	0.64	0.68
CVA model	English	0.38	0.27	0.23	0.29
	Maths	0.70	0.77	0.58	0.68
	Science	0.67	0.63	0.62	0.64

Table 7: Level 3 (school level) random effects by pupil characteristics

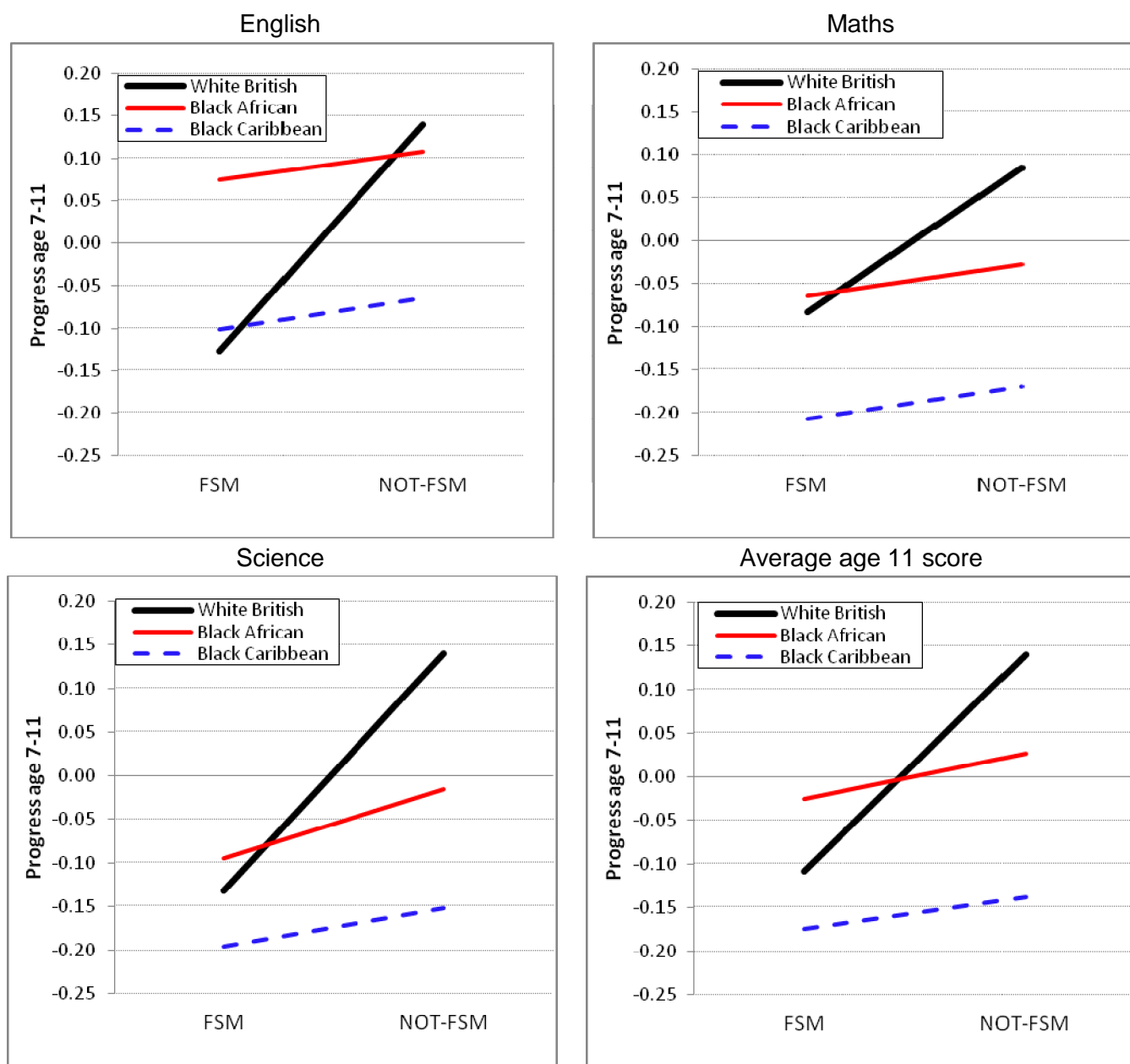
Outcome & variable	Est.	SE		SD	Level 1 coeff.	5th centile school	95th centile school
English							
Age 7 score	0.006	0.002	*	0.077	0.57	0.41	0.72
FSM	0.004	0.003		0.063	-0.26	-0.39	-0.13
Girl	0.000	0.002		0.010	0.17	0.15	0.19
Black Caribbean	0.002	0.003		0.045	-0.20	-0.29	-0.11
Maths							
Age 7 score	0.006	0.002	*	0.077	0.62	0.46	0.77
FSM	0.001	0.003		0.025	-0.17	-0.22	-0.11
Girl	0.003	0.003		0.055	-0.21	-0.32	-0.10
Black Caribbean	0.003	0.004		0.055	-0.26	-0.37	-0.15
Science							
Age 7 score	0.007	0.002	*	0.084	0.54	0.37	0.71
FSM	0.002	0.003		0.045	-0.26	-0.35	-0.17
Girl	0.002	0.003		0.045	-0.11	-0.20	-0.02
Black Caribbean	0.002	0.004		0.045	-0.29	-0.38	-0.20
Age 11 avg. score							
Age 7 score	0.007	0.002	*	0.084	0.63	0.46	0.80
FSM	0.002	0.002		0.045	-0.25	-0.33	-0.16
Girl	0.001	0.002		0.032	-0.06	-0.13	0.00
Black Caribbean	0.001	0.003		0.032	-0.28	-0.34	-0.21

Notes: * indicates $p < .05$.

Table 8: Correlations between school residuals for different models

Subject	Correlation	Raw vs. VA	Raw vs. CVA	VA vs. CVA
English	Lowest in any year	0.84	0.75	0.96
	highest in any year	0.85	0.82	0.97
	Average across years	0.84	0.78	0.96
Maths	Lowest in any year	0.88	0.82	0.94
	Highest in any year	0.90	0.83	0.96
	Average across years	0.89	0.82	0.95
Science	Lowest in any year	0.89	0.84	0.97
	highest in any year	0.91	0.88	0.98
	Average across years	0.91	0.86	0.97

Figure 1: Pupil progress age 7-11 by ethnic group and entitlement to FSM for English, maths science and average age 11 score



*Notes: estimates of progress based on regression coefficients for FSM, the three main ethnic groups and the interactions between ethnic group * FSM. A constant value has been added to centre results at the mid-point of the FSM gap for White British pupils. Coefficients are calculated at the average age 7 score and the base level of all other variables listed in table 2.*

Figure 2: CVA school residuals with 95% confidence intervals for the 57 schools for each subject and for average age 11 score

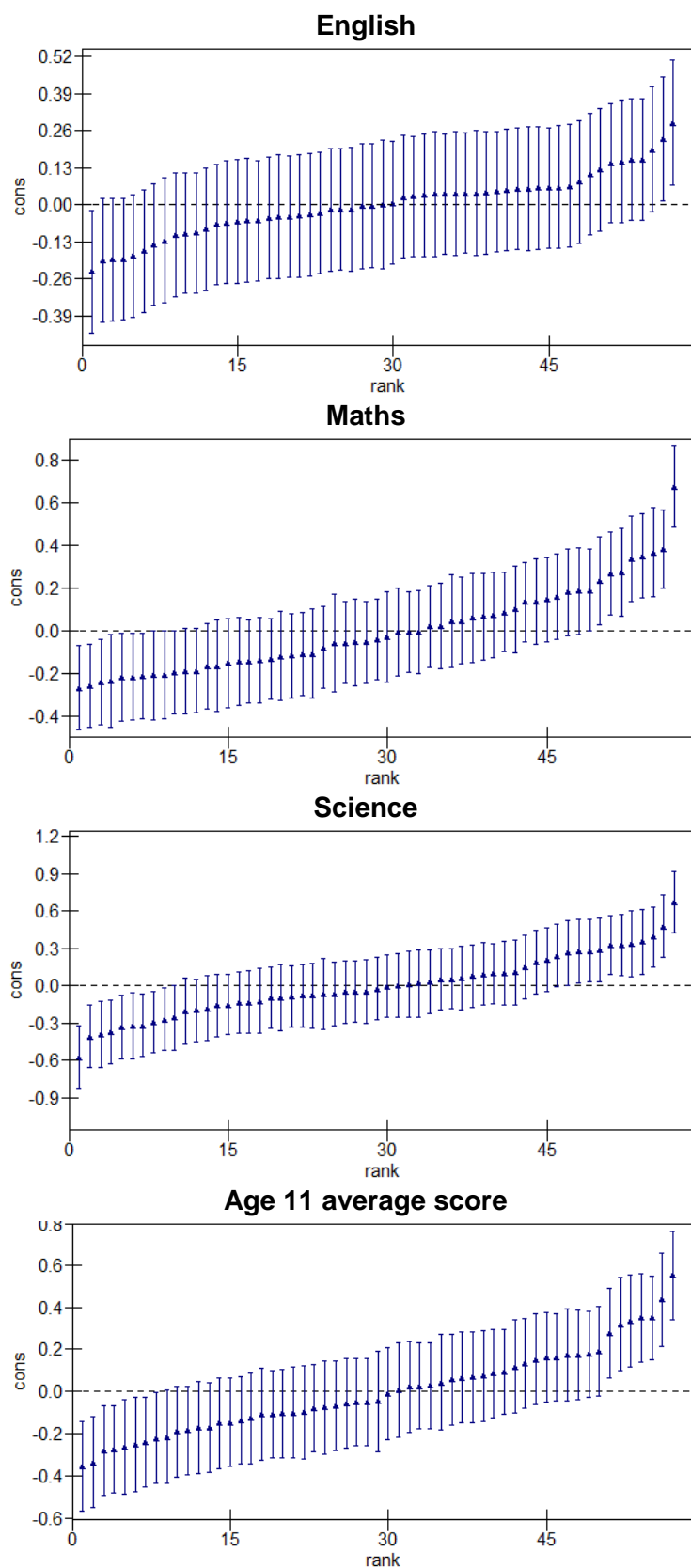


Figure 3: Correlation between school CVA residuals for English and maths 2006 ($r=0.66$)

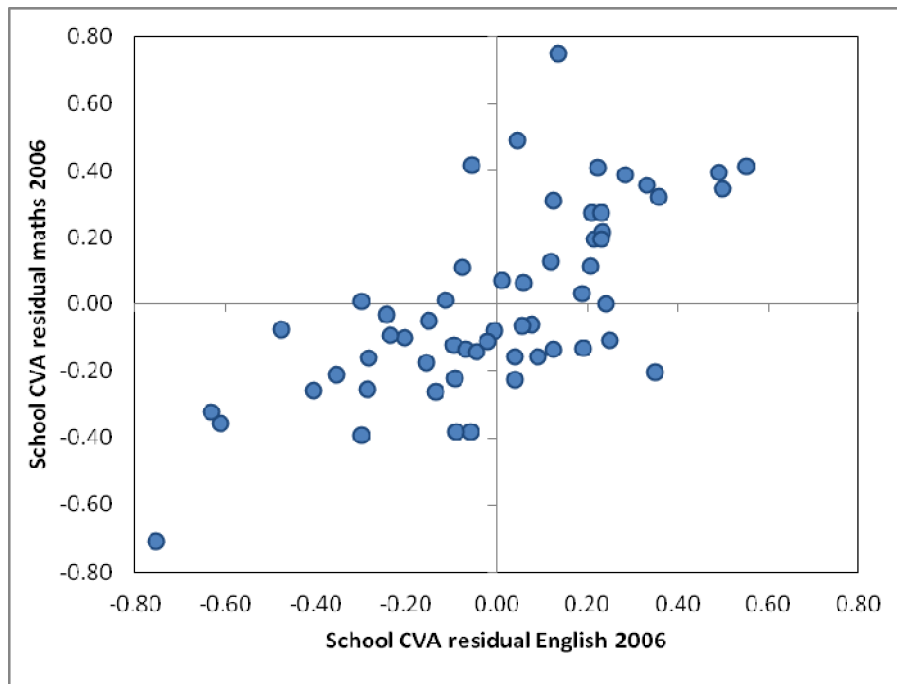


Figure 4: Correlation between school CVA residuals for mathematics 2004 vs. 2006 ($r=0.58$)

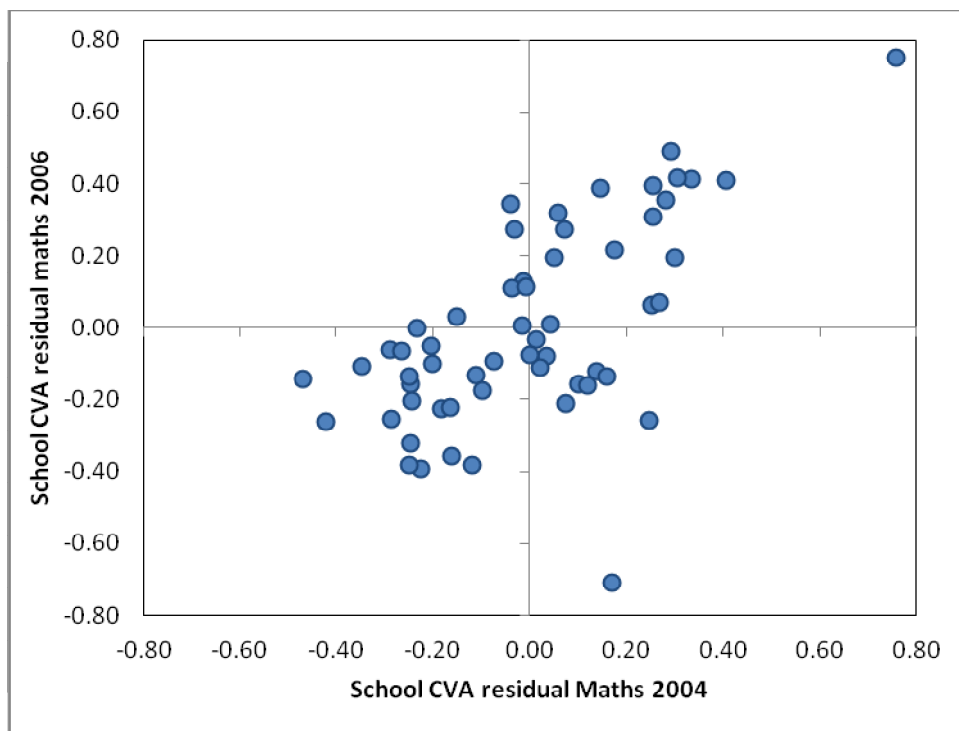


Figure 5: School CVA residuals (average age 11 score) at different levels of prior age 7 score (r low vs high= 0.79)

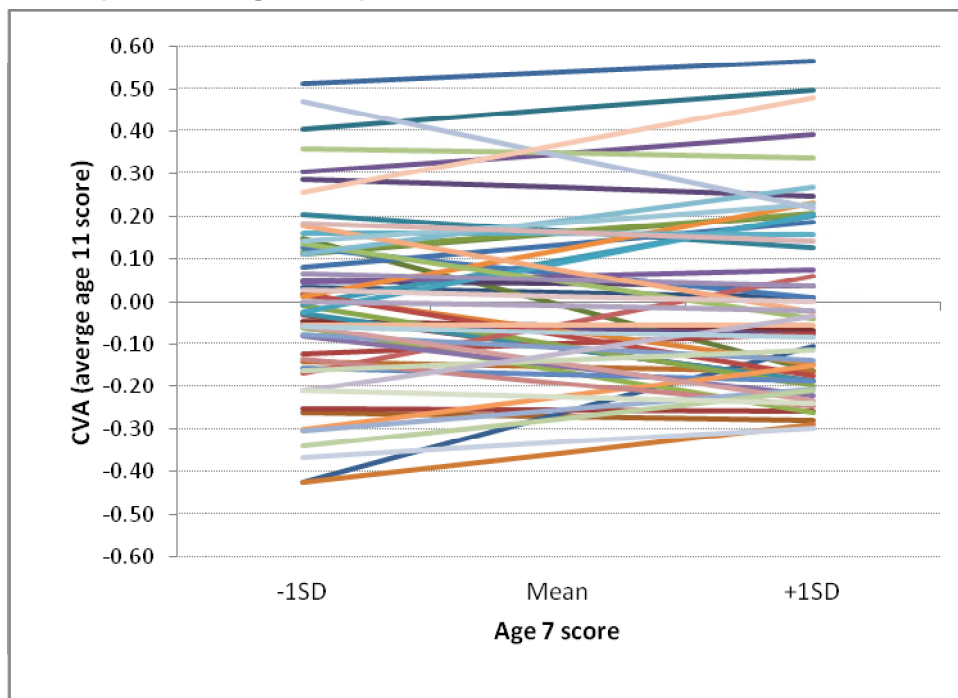


Figure 6: Correlation between school CVA residuals (average age 11 score) for pupils entitled and not entitled to FSM ($r=0.98$)

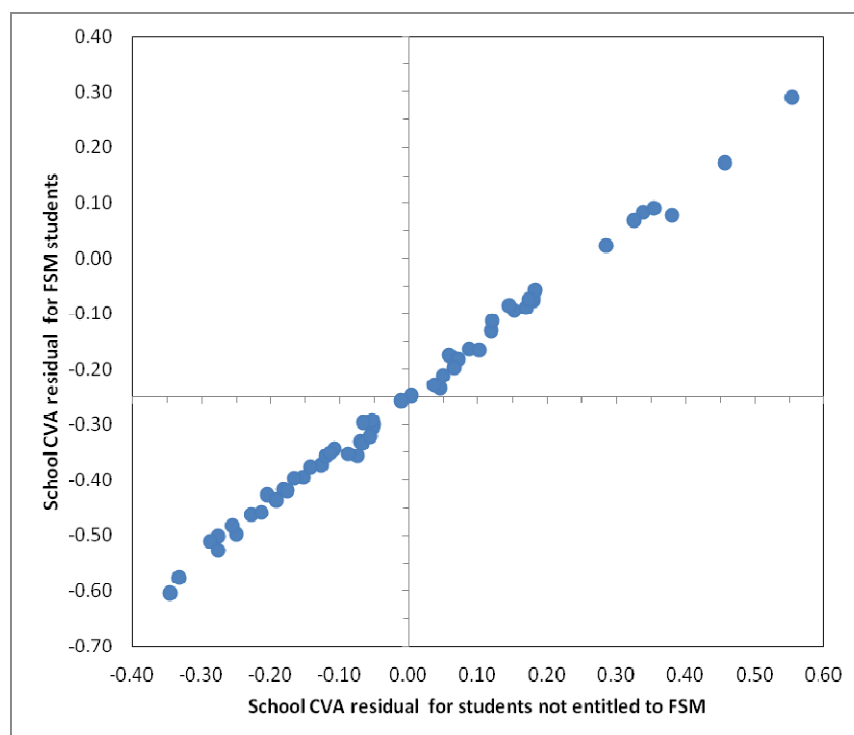


Figure 7: Correlation between school CVA residuals (average age 11 score) for White British and Black Caribbean pupils ($r=0.91$)

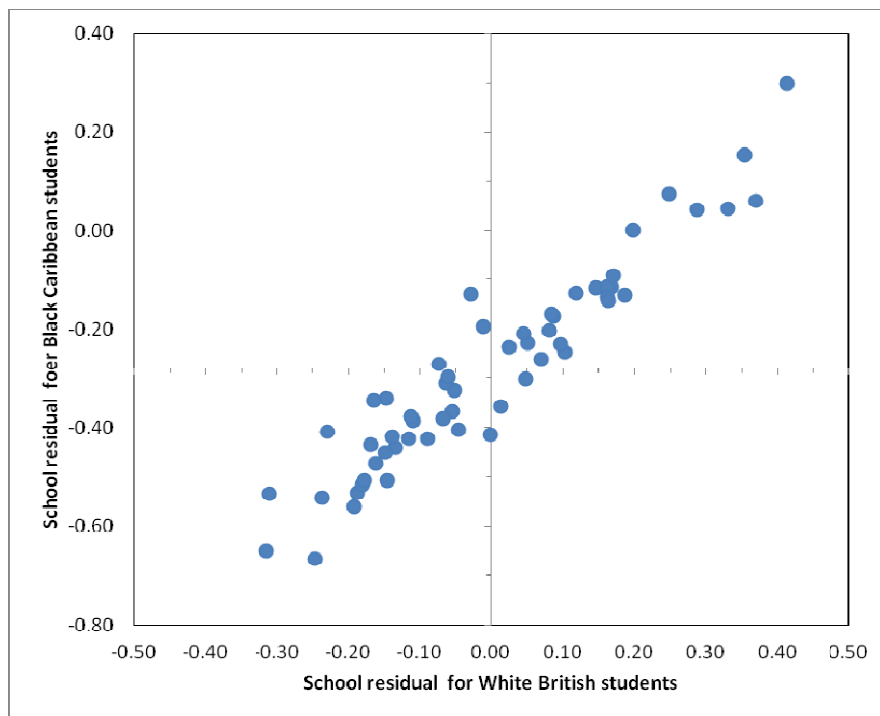


Figure 8: Correlation between school raw and CVA scores for English in 2005 ($r=0.78$)

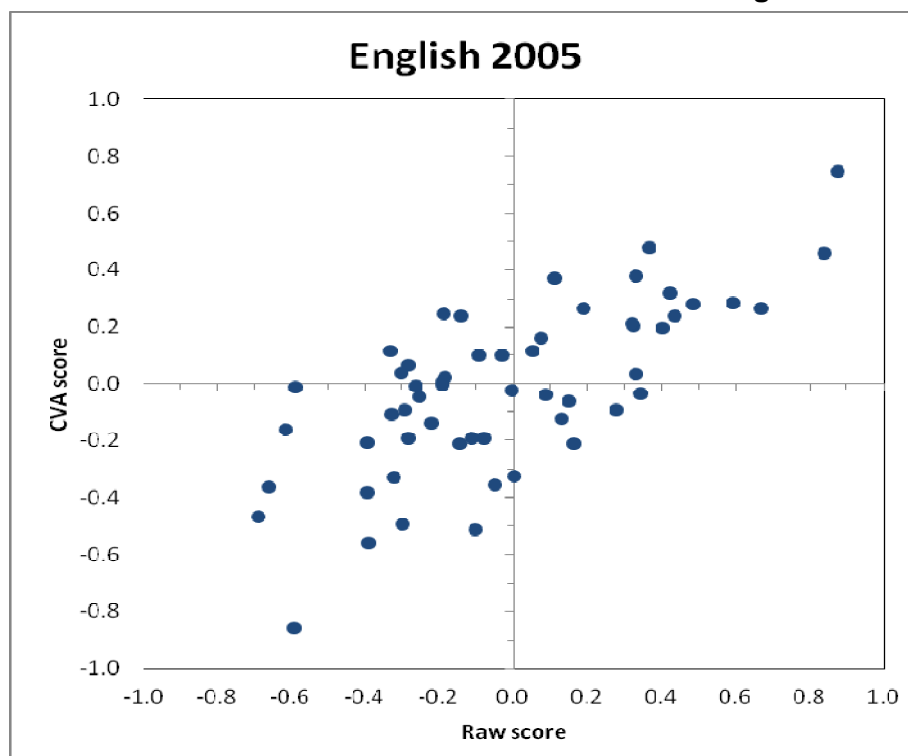


Figure 9: Correlation between school VA and CVA scores for Mathematics in 2004 ($r=0.94$)

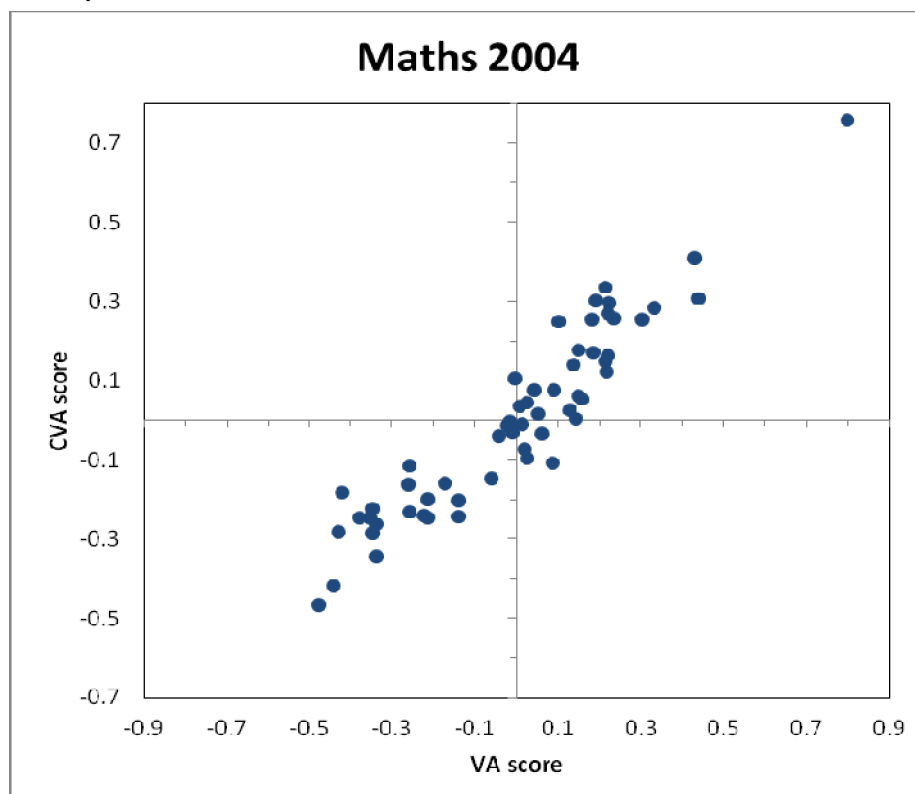
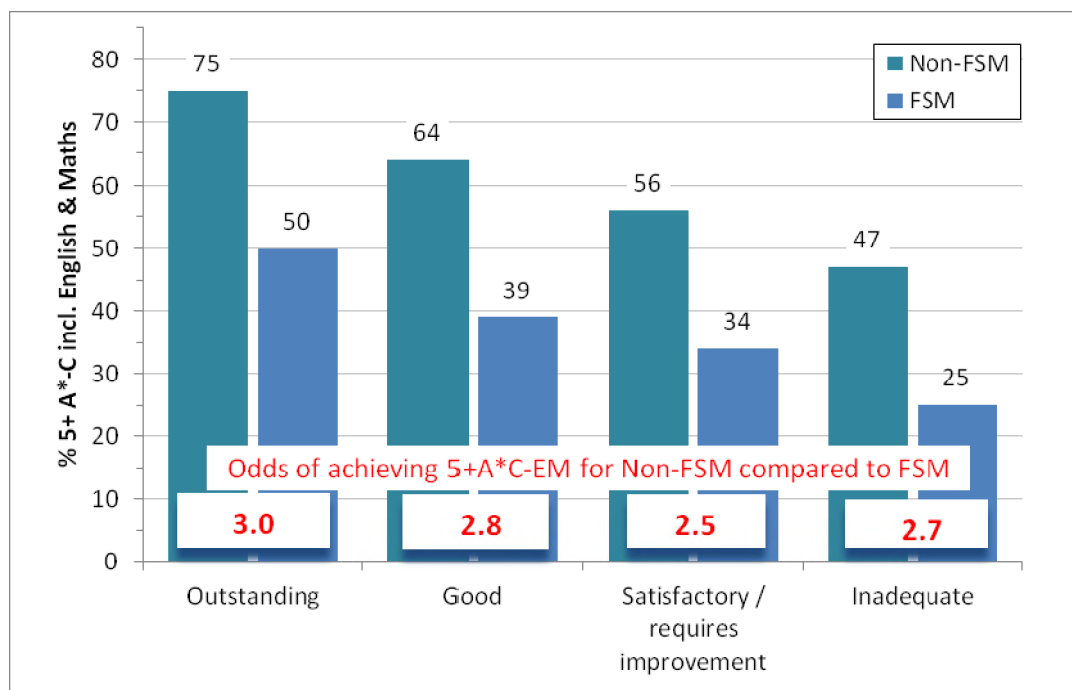


Figure 10: Achievement of FSM and non-FSM pupils by the OFSTED judgment of the schools' overall effectiveness



Notes: Data sourced from Ofsted (2013), p53.

FOOTNOTES

- i. *Sponsored academies* are low performing previously Local Authority (LA) maintained schools compelled to convert to academy status. These schools are directly funded by the Secretary of State and run by an external sponsor through a charitable company called an academy trust. Most sponsored academies are in chains of three or more academies run by a single sponsor. *Converter academies* are previously outstanding or good schools typically with low numbers of disadvantaged pupils who have opted out of LA control. In December 2014 there were 4,344 open academies, including over half of all secondary schools in England. See House of Commons (2015) for further details.
- ii. Data for a small number of pupils attending special schools were excluded from the sample.
- iii. This was the case when the current data were collected, although since 2010 the science tests are no longer universally administered, being completed only by a representative sample of 5% of primary schools each year for national monitoring purposes.
- iv. Specifically combined were: Bangladeshi & Pakistani; Indian, Vietnamese, Chinese and Other Asian; and all mixed heritage groups (other than Mixed White and Black Caribbean). A small number of pupils where ethnicity was not obtained or refused (n=48) were combined with any other group.
- v. Statistical significance tests are reported here because I believe that the Year 6 children observed in this study are just a sample of children in these schools among many possible samples. I am interested here in what has been termed model-based rather than design-based inference (see Plewis & Fielding, 2003; Goldstein & Noden, 2004, p441). However since the dependent variables have been normalised all associations with explanatory variables are expressed as effect sizes making it possible to evaluate the substantive as well as the statistical significance of all associations.
- vi. The three way interaction between ethnic group, gender and FSM, and the two way interactions between gender and FSM and between ethnicity and gender were also tested but were not significant and were eliminated.
- vii. These effects are calculated directly from the regression coefficients. White British pupils not entitled to FSM are the reference category. For ethnic minority pupils not on FSM the scores are given by the ethnic coefficients in Table 2. For pupils on FSM the White British score is given directly by the FSM coefficient, while the scores for each ethnic group are found by adding the FSM, ethnic and relevant ethnic*FSM coefficients. The scores are adjusted for all other variables included in the model. A constant has been added to each equation to centre the results at the mid-point of the White British FSM gap.
- viii. Tests of ethnic contrasts among pupils entitled to FSM were made by reversing the reference group to be FSM (rather than non-FSM) pupils (see Jaccard & Turrisi, 2003).
- ix. Following Tymms (2004) the effect size for a continuous variable is calculated as the $(\text{coeff} * 2 \text{ SD}) / \text{the pupil level SD}$ (the latter being equal to 1 because of the normal score transformation). Specifically, given cohort %FSM had a mean of 38 and SD of 15, the effect size equal $(0.006 * 30) / 1 = -0.18 \text{ SD}$. This indexes the difference between a school with 23% FSM and a school with 53% FSM.
- x. For example for English in the null model the within-school cohort variance is $.051 / (.051 + .102)$ or 33% of the total between school variation, and the average correlation between cohorts over time is $.102 / (.102 + .051)$ or 0.67. See Leckie, 2013, p24-26.

-
- ^{xi} . The correlations over time for CVA scores for maths and science reported by Melhuish et. al. (2006) are lower than reported here, possibly because they used NC levels rather than test marks.
 - ^{xii} . The outcomes for the multivariate and univariate models were actually identical, possibly because there is very little missing data, just 17 pupils missing either an English, mathematics or science score.
 - ^{xiii} . Given there were no significant effects when each parameter was tested alone, attempts to assess random variation in all parameters simultaneously often failed to converge or gave estimates that were so close to zero as to be set to zero.
 - ^{xiv} . This is achieved by no longer allowing the intercept term (Cons) to vary randomly at the school level and instead including an additional coefficient for NOT-FSM as well as the coefficient for FSM, with both coefficients fixed and allowed to vary at the school level.
 - ^{xvi} . OFSTED judgments of a schools' effectiveness are frequently used by politicians and others as measures of school quality, but it should be noted this is problematic since OFSTED judgements are heavily influenced by the published examination results achieved by the school. Also it should be noted it is no longer possible to replicate the analysis reported here since from September 2015 one of the criteria necessary for OFSTED to judge a school 'Outstanding' is that it has a small FSM gap.
 - ^{xvii} . The data published since 2012 are for 'disadvantaged pupils' which is all those who have been entitled to a FSM anytime in the last six years and Children Looked After (CLA) by the Local Authority.