Institute of Education

The Short Run Impact of the Building Schools for the Future Programme on Attainment at Key Stage 4

Dave Thomson

Department of Social Science
Working Paper No. 16-07

April 2016

Disclaimer

Department of Quantitative Social Science, UCL Institute of Education, University College London, 20 Bedford Way, London WC1H 0AL, UK

The Short Run Impact of the Building Schools for the Future Programme on Attainment at Key Stage 4

Dave Thomson[1]

Abstract

Building Schools for the Future (BSF) was a £55 billion, 15 year programme to rebuild or renovate all secondary schools in England that was cancelled after 6 years. By comparing pupil attainment at schools whose projects were completed to pupil attainment at schools whose projects were cancelled, the effects of new school buildings on pupil attainment are estimated. A number of different estimation methods are used, including linear regression, conditional difference-in-differences (with and without propensity score matching), and 'within-between' random-effects regression. Results from the various models are broadly similar and show that new school buildings have no effect on pupil attainment, at least in the short-term. Given that the stated aim of BSF was educational transformation, such outcomes represent poor value for money in the short term.

JEL codes: I28

Keywords: School buildings; Pupil attainment; Building schools for the future programme

Contact Details: Dave Thomson (dave.thomson@fft.org.uk), Education Datalab

[1] Education Datalab, FFT
dave.thomson@fft.org.uk

1. Introduction

Whilst previous research has examined the impact of individual elements of school design, such as lighting and acoustics, on pupil attainment (see Higgins et al, 2005 for a review), Edgerton et al (2010) contend that these studies do not "capture the complexity of the 'whole' school experience." When attempted, research into the impact of school buildings in England has tended to be small scale (Leiringer & Cardoza, 2009) or assessed in relation to pupil and teacher perceptions (Edgerton et al, 2011).

Nonetheless, there is some promising evidence that school buildings and design have a positive effect on attainment. In England, Barrett et al (2014) find relatively large effects in their study of the impact of classroom design on pupil progress in 27 primary schools in 3 local authority areas. They identify seven design elements, such as light and air quality, that together explain 16% of the variation in the amount of progress made by pupils in a single year. Meanwhile in the USA, Neilson and Zimmerman (2011) find small but significant impacts of school buildings on test scores, house prices and school enrolment in poor urban areas. Exploiting a natural experiment, they observe flat trends in reading scores prior to construction which then turn upwards for at least the next six years. Overall, they calculate an increase of 0.21 standard deviations in reading for pupils receiving the average treatment intensity. By contrast, Martorell et al (2015) find little evidence of the causal impact of capital expenditure on student achievement in school districts in Texas.

Building Schools for the Future (BSF) was a flagship policy of the new Labour government which served from 1997 to 2010. It aimed to transform education, and in doing so raise attainment, by renovating all 3,500 secondary schools in England (PfS/4ps, 2008). However, the James Review of 2011 would later scathingly dismiss the aim of educational transformation as 'laudable, but undefined' (James, 2011). The Programme was eventually halted by the coalition government which came to power in May 2010. By this time, some schools had been renovated and many others were in progress. However, those which had not reached 'financial close' in the procurement process were cancelled.

The decision to cancel BSF projects presents not only a rare opportunity to investigate the effect of a major school building programme on educational attainment in England but also more generally the effect of new school buildings. These effects are estimated by comparing renovated schools to those whose projects were cancelled. This reduces but does not wholly obviate the selection problem that would be inherent in a naïve evaluation comparing renovated schools to all other schools.

A number of modelling approaches are used to estimate the effect of the programme, all of which produce substantively similar results. Firstly, the 'treatment vs comparison' specification is a straightforward regression model which compares rebuilt schools ('treatment') with those that were in the process of being rebuilt or which had been cancelled ('comparison'). Secondly, the 'difference-in-differences' specification evaluates trends in attainment post-treatment in light of trends in attainment among 'comparison' schools.

As schools in the earlier waves of BSF were more likely to have been rebuilt before the programme was cancelled, propensity score matching (Rubin, 1977) is used to balance the sample in terms of the criteria by which schools were selected for BSF. These were low academic attainment and high deprivation at local authority level. In practice however, balancing the sample using the derived weights makes little difference to the impact estimates.

The models are applied to datasets covering ten academic years. Both pupil-level and aggregate school-level datasets are used. The latter is effectively a ten-year strongly

balanced panel. A range of covariates known to have an effect on attainment, such as prior attainment, deprivation and gender are included. As a number of schools have merged, closed or opened over the ten year period, schools are indexed by their latest identifier and only schools with the full ten years of data are considered.

As is the case with all non-experimental retrospective evaluation, the models are predicated on a number of identifying assumptions. Firstly, the stable unit treatment value assumption (Rubin, 1980) that the treatment affects only the treated; in this case that BSF had no effect on schools whose projects were cancelled. Secondly, the common trends assumption (Blundell et al, 2004) that trends in attainment among the 'treatment' and 'comparison' schools would have been the same in the absence of the programme. Thirdly, the 'treatment vs comparison' specification assumes that the effects of unobserved variables, such as levels of funding, are invariant over the ten years covered by the datasets. Finally, the 'difference-in-differences' specification assumes that the attainment outcomes of 'treatment' and 'comparison' schools are independent conditional on observed covariates.

Overall, no significant impact of new school buildings is found. However, allowing for heterogeneous impacts reveals a small, lagged effect in 2012/13 among the earliest cohort of BSF schools, those which were completed by December 2010. The application of a multilevel modelling framework to the 'difference-in-differences' specification reveals significant variation in the effect of BSF among 'treatment' schools.

The implications of this research are twofold. First and foremost, it provides evidence of the impact of new school buildings for use in cost/ benefit analysis. Such estimates were unavailable to those planning BSF projects who were thus unable to set realistic ambitions. Secondly, it raises the question of whether replacing old school buildings with new has a significant impact on some types of school, such as those with low levels of attainment. In their analysis of the Academies Programme in England, Machin & Vernoit (2011) find significant impacts on attainment among the first wave of Academies, a new type of state-funded school outside of local authority control introduced in the early part of the century. These were not apparent for later waves once the Programme had been accelerated. The first Academies, which opened between 2002/03 and 2006/07, tended to replace low-attaining 'failing' schools, often in deprived areas. In most but not all cases, the previous school was demolished and new facilities were constructed. It could be the case that the change of buildings rather than the change of governance engendered significant improvements in attainment.

The structure of the rest of the paper is as follows. Section 2 provides more detail on the Building Schools for the Future programme. Section 3 describes the datasets created. The estimation methods and identifying assumptions are specified in section 4. Results are presented in section 5 and section 6 concludes with suggestions further opportunities for research in this area.

2. Building Schools for the Future

Announced in February 2003 and commencing the following year, Building Schools for the Future (BSF) was a £55 billion investment programme to renovate all 3,500 secondary (high) school buildings in England over the period 2005 to 2020 (PriceWaterhouseCoopers, 2010; National Audit Office, 2009). It planned to rebuild around half of schools, remodel (i.e. partially rebuild) around 35% and refurbish the remainder. Refurbishment included investment in Information and Communications Technology (ICT).

BSF was delivered through local education partnerships (LEPs) between local authorities, private sector contractors, lenders, the Department for Education and Partnerships UK, a joint venture between government and the private sector to fund public projects through private finance (4ps & Partnerships for Schools, 2008). Early evaluations of the programme were critical of this delivery model which was characterised by delay, high costs and poor project management (National Audit Office, ibid).

The costs of the programme were justified on the assumption that new school buildings would have transformative effects upon educational attainment and social justice (Mahony and Hextall, 2011). Indeed, each local authority was required in its formal submission to draw up a 'Strategy for Change', detailing how such transformation would arise from investment in the school estate. Mahony et al (2011) describe how the "the stated purposes and roles of BSF…shifted and (were) in some cases contradictory" during the lifetime of the project. Although published without references to supporting evidence, official BSF documentation (PfS & 4ps, 2008) asserted that transformation be achieved by designing learning environments that would:

- Facilitate personalised learning;
- Imbue a sense of safety and security;
- Inspire learning through the use of computer technology; and
- Extend the traditional school day, offering extra-curricular learning and leisure opportunities.

The four themes above could be considered as direct pedagogical effects of school buildings on attainment. One might also add the design of learning spaces that enable behaviour to be better maintained (Foucault, 1975). Additionally, new buildings may have motivational effects upon teachers and pupils. Neilson and Zimmerman (2011) investigated both by means of a small-scale qualitative survey of 22 school principals before, during and after school construction. This revealed large effects on student and pupil motivation and increased parental engagement. These were found to be at least as important as the direct effects of improved buildings on pedagogy.

The programme was to be rolled out in fifteen waves, with the first tranche of funding distributed to local authorities in 2005/6 (National Audit Office, op. cit.). Small local authorities were allocated to a single wave whilst projects in larger authorities were scheduled in multiple waves. The initial allocation of waves prioritised local authorities largely on the basis of secondary school attainment (the percentage of pupils achieving 5 or more GCSEs at grades A*-C) and deprivation (the percentage of pupils eligible for free school meals). In addition, the 25 local authorities destined to join the programme last of all had the opportunity during early waves to bid for funding to each rebuild a single school in the greatest need, known as One School Pathfinders.

It was intended that it would take local authorities around two and a half years to begin construction once they had formally joined the programme. During this time, a governance structure was to be established, the 'Strategy for Change' and 'Outline Business Case' completed, tender documentation prepared and a preferred bidder selected (PfS/4ps, 2008). Rather than procure each project in a separate tender exercise, the preferred bidder would rebuild all selected schools within a local authority.

As the James Review later noted, there were a number of concerns about the quality of some of the school designs. Consequently, in 2007 the Commission for Architecture and the Built Environment (CABE) was commissioned by the then Department for Education and

Skills (DfES) to introduce Design Review Panels, which rated designs on a 5 point scale ranging from ‘excellent’ to ‘poor’.

> These panels judged all of the 63 early BSF designs, and these were shown to be inadequate even after several stages of re-design. Many of them were at too late a stage in the project process to be stopped, and were nevertheless built (James, 2011, 2.39).

This initial rating exercise led to the creation of minimum design standards (MDS) introduced in summer 2009. School designs were subsequently rated on a 1-4 scale from ‘very good’ to ‘unsatisfactory’.

In May 2010, a coalition government formed by the Conservatives and Liberal Democrats replaced the Labour government which had established BSF. In June 2011, Michael Gove, the incumbent Secretary of State for Education, announced plans to end BSF as part of the new Government’s austerity measures. Whilst some planned projects remained unaffected, those that had not reached ‘financial close’ were cancelled with immediate effect, save for a small number that would continue under a different capital funding stream- the Academies Programme- and which would also lead to changes in the governance of the schools affected (see Machin and Vernoit, 2011).

By March 2011, almost £9 billion had been spent of which 60% was funded through the Private Finance Initiative (PFI) and the remainder through conventional capital funding (James, 2011). Under this arrangement, private investors fund the design, build and operation of buildings which are then leased back from the Government. The cost of capital for a typical PFI project is currently over 8%, double the cost of conventional government borrowing, which has led to the value for money of PFI projects being repeatedly called into question (House of Commons Treasury Committee, 2011). The costs of rebuilding schools under BSF will therefore be borne by taxpayers for many years to come.

3. Data

Although initially beset by errors, the Department for Education published the status of each project in each local authority at the time of its announcement to cease BSF[i]. Status took one of five values:

Open	Projects for which a planned new building, refurbishment, extension or ICT had been delivered. A month and year of opening was also present in the majority of cases.
Stopped	Projects which had not yet reached ‘Financial Close’ (other than the ‘sample’ schools in projects which had passed the ‘Close of Dialogue’ point) or which were Academy Framework projects, which had not achieved ‘Financial Close’, and where there was no funding agreement in place and the Academy was not open or about to open; or which were to have been in a repeat wave of investment, but which had not received approval prior to 1 January 2010.
Unaffected	Projects which had reached ‘Financial Close’; or were within a repeat wave of investment
Academy- for discussion	Academies where the building projects had not reached financial close, but which were either already open, had a signed Funding Agreement, or

	were due open in the next academic year or were making completely new provision.
Sample[ii]- for discussion	Schools which were to have been 'sample' schools in projects which were at an advanced stage in the procurement process (that is, they had passed the 'Close of Dialogue') but which had not yet reached 'Financial Close'.

Desk research was undertaken to identify the month and year of subsequent opening for schools which were listed as 'unaffected' or 'for discussion'. In most cases, this was the day a new or remodelled school opened to pupils. In cases where a project consisted of a number of new buildings, the month and year in which the first building opened to pupils was found. Projects consisting of refurbishment only, including schools in which only investment in ICT equipment was made, were also identified.

Excluding projects identified as refurbishment/ ICT only, those relating to special schools and pupil referral units and those which established new provision, 485 secondary schools were categorised as rebuilt or remodelled (Table 1) under BSF. A further 467 projects were stopped and 99 were listed as stopped under BSF but which would be funded instead under the Academies Programme. Projects in local authorities that joined BSF in later waves (particularly 6 and 7) were more likely to be stopped although some projects in LAs that joined BSF earlier were also stopped. These tended to be in local authorities in which projects were implemented in multiple phases. Additionally, some 'one school pathfinder' projects were completed in local authorities destined to join BSF in wave 8 or later.

Table 1: Number of BSF rebuilding and remodelling projects by wave in which local authorities first joined the programme (mainstream schools only)

Wave	Rebuilt/ remodelled	Stopped	Stopped- for Academy
Pilot	47	19	2
1	98	38	0
2	78	33	10
3	71	34	6
4	50	49	6
5	38	49	4
6a	19	113	18
6b	5	36	8
7	46	90	28
8 to 15	33	6	17
Total	485	467	99

Of those that were stopped, around 60 were subsequently due to be funded from the Priority Schools Building Programme (PSPB), a slimmed-down successor to BSF announced in May 2012. Some of the projects shown as rebuilt or remodelled were still in progress at the time of writing, with the final projects due to complete in the 2015/16 academic year.

Including those classified as 'Stopped- for Academy', 584 schools were rebuilt or remodelled. In Table 2, five cohorts of BSF schools are identified based on the year in which they opened as follows:

- Those still under construction at the end of December 2013
- Those completed in the calendar year 2013
- Those completed in the calendar year 2012
- Those completed in the calendar year 2011
- Those completed in the calendar year 2010 or earlier

Table 2: Number of completed BSF projects by year of opening and wave in which local authorities joined the programme (mainstream schools only)

	BSF Cohort				
Wave	>2013	2013	2012	2011	<=2010
Pilot	2	5	11	12	19
1	8	8	14	12	56
2	4	22	23	18	21
3	13	14	14	23	13
4	11	11	26	5	3
5	9	14	12	4	3
6a	8	14	6	3	6
6b	1	7	3	1	1
7	7	25	20	10	12
8 to 15	3	10	8	10	19
Total	66	130	137	98	153

Pupils in state-funded schools in England follow a National Curriculum, divided into four 'Key Stages' from statutory admission at age 5 to the end of compulsory schooling at age 16. Key Stage 4 covers the final 2 years of compulsory schooling. Data on pupils attending the schools shown in Table 1 for the years 2002/3 to 2012/13 were extracted from the National Pupil Database (NPD; DfE, 2012). This is a longitudinal database of pupil attainment from age 7 (Key Stage 1) through to the end of Key Stage 4 (at age 16) and beyond. In addition, data on attainment can be linked (via a student identifier) to data on school enrolments and pupil characteristics (such as gender, ethnicity, special educational needs) collected from the School Census[iii]. As data on pupils not entered for public examinations (GCSEs) or equivalent qualifications at Key Stage 4 was not available for previous years, 2003/4 is the year in which our window of analysis commences. This was also the first year in which other qualifications equivalent to GCSE, often referred to as Section 96 qualifications, were counted in School Performance Tables[iv].

Models are estimated for two Key Stage 4 outcomes. Firstly, the percentage of pupils achieving 5 or more A*-C grades (or equivalent) including English and mathematics and secondly, the mean grade in GCSE examinations. As the most-widely known and used headline indicator of secondary school attainment, the former indicator is considered 'high stakes' and has tended to be used to drive school-level interventions, particularly for schools below 'floor targets'. Consequently, it has tended to produce an incentive for some schools, particularly those at risk of falling below the floor target, to focus their efforts on pupils at the C/D borderline (Goldstein & Foley, 2011) at the expense of other pupils. By contrast, mean GCSE grade is an unpublished, 'low stakes' indicator. It is attractive in that it measures attainment in a set of qualifications that have been both available and widely entered over the whole of the time series from 2003/4 to 2012/13, avoiding the issue of the equivalence of other types of qualification (Wolf, 2011). As published indicators have tended to change in

definition over time, the most recent definitions of both indicators have been retrospectively applied to examination-level data[v] and subsequently aggregated to both pupil and school level.

Outcomes are modelled conditional on a number of time-varying covariates including pupil prior attainment (test scores at the end of primary school) and pupil characteristics including gender, deprivation (free school meal eligibility), ethnicity and special educational needs. These are standard controls used in 'value added' models of pupil progress (see Jenkins et al, 2006). For school-level models, these variables are aggregated from pupil level data.

The impact of BSF on school composition is tested through an additional set of difference-in-differences models (equations 3.1 to 3.3) with the percentage of year 7 pupils known to be eligible for free school meals (FSM) the dependent variable but an empty set of school-level covariates *x*. FSM eligibility is a well-used measure of economic disadvantage that is consistently available in administrative data over a number of years (Hobbs & Vignoles, 2009). Year 7 marks the transition from primary to secondary school. Admission to secondary school is based upon a process of application and allocation (see West et al, 2009), with many popular schools receiving many more applications than places available (over-subscription). A change in school composition may be the result of a local change in school popularity.

As a result of mergers, closures (including when a maintained school closes to make way for a sponsored Academy), amalgamations and changes of governance, school identifiers change over time. A history of such changes affecting the schools in Table 1 over the period 2002/3 to 2012/13 was created using the Schools Database (SCDB). This allows, where appropriate, comparison between an institution and its predecessor(s). In some parts of the country (e.g. Burnley), two schools were closed and replaced with a single new school built under BSF. By contrast, Westminster Community School was replaced by two Academies, Paddington and Westminster. Due to these changes, schools are indexed by their latest school identifier in the analysis that follows.[2]

4. Estimation Methods

A number of non-experimental models are used to estimate the impact of 'treatment' under BSF in relation to a 'comparison' set of schools consisting of those whose projects were stopped. This helps to reduce the selection problems that would be inherent in, for example, comparing BSF schools with all other schools.

Two approaches to modelling school-level data are taken. Firstly, a simple between-schools comparison of pupil attainment in the treatment schools to the 'comparison' set, controlling for time-varying school-level covariates (treatment vs comparison). Secondly, using a similar differences-in-differences methodology adopted by Machin and Vernoit (2011) in their analysis of the impact of the Academies programme. Results based on this approach are compared to results from analogous models using pupil-level data.

The basic model specification for the treatment vs comparison model is shown in 1.1. It is estimated separately for each outcome for each year within the window of analysis, weighted by pupil numbers. X_{js} represents a set of time-varying covariates for each school. μ_s is the error term for each school s. $treatment_{cs}$ is a Boolean flag that identifies schools that were either remodelled or rebuilt under BSF or for which a project was still under construction. Five cohorts of schools c are identified based on year of opening.

[2] For example, 2 schools which merged in 2010 will be identified as the merged school in years prior to 2010

$$y_s = \sum_{c=1}^{5} \delta_c \, treatment_{cs} + \sum_{j=0}^{J} \beta_j \, X_{js} + \mu_s \qquad (1.1)$$

The basic model specification for the school-level difference-in-difference model is shown in 2.1. This involves creating a set of fixed school effects for each school (α_s) and a set of fixed year effects for each year (α_t) relative to the first year of the analytical window (2003/4). X_{jst} represents a set of time-varying covariates for each school. μ_{st} is the error term for each school s in year t. Finally, $policyon_{st}$is a boolean flag set to 'true' the year in which a school BSF project is completed (i.e. begins to be used for teaching) and all subsequent years. Schools are weighted by pupil numbers. The value of δ is therefore a difference-in-difference (DID) estimate of the impact of BSF.

$$y_{st} = \alpha_s + \alpha_t + \delta policyon_{st} + \sum_{j=0}^{J} \beta_j \, X_{jst} + \mu_{st} \qquad (2.1)$$

The equation in 2.1 would yield a 'one-off' impact of BSF that would be consistent across all years. To allow for heterogeneous impacts conditional on the academic year in which BSF projects were completed, three 'cohorts' *c* of schools are created based upon academic year of opening, with 2012 the latest cohort and all schools which opened in 2010 or earlier the first cohort. $policyon$ is set to zero for 'treatment' schools which had not opened by the end of December 2012. This models yields impact estimates for each cohort.

$$y_{st} = \alpha_s + \alpha_t + \sum_{c=1}^{3} \delta_c \, policyon_{cst} + \sum_{j=0}^{J} \beta_j \, X_{jst} + \mu_{st} \qquad (2.2)$$

Finally, to allow for differential effects over time, the $policyon$ variable for each cohort in 1.2 is interacted with time. This model yields an effect for each cohort (as in 1.2) for each year.

$$y_{st} = \alpha_s + \alpha_t + \sum_{c=1}^{3} \delta_{ct} \, policyon_{cst} * \alpha_t + \sum_{j=0}^{J} \beta_j \, X_{jst} + \mu_{st} \qquad (2.3)$$

Analogous pupil level models are specified in 3.1 to 3.3. As with the school-level models, there is a set of fixed school effects for each school (α_s) and a set of fixed year effects for each year (α_t). X_j represents a set of prior attainment covariates for each pupil p. μ_{pst} is the error term for each pupil p in school s in year t. Model 2.2 adds in heterogeneous impacts for each cohort of schools and 2.3 allows these impacts to vary over time.

$$y = \alpha_s + \alpha_t + \delta policyon_{st} + \sum_{j=0}^{J} \beta_j \, X_{jpst} + \mu_{pst} \qquad (3.1)$$

$$y_{st} = \alpha_s + \alpha_t + \sum_{c=1}^{3} \delta_c \, policyon_{cst} + \sum_{j=0}^{J} \beta_j \, X_{jpst} + \mu_{pst} \qquad (3.2)$$

$$y_{st} = \alpha_s + \alpha_t + \sum_{c=1}^{3} \delta_c \, policyon_{cst} * \alpha_t + \sum_{j=0}^{J} \beta_j \, X_{jpst} + \mu_{pst} \qquad (3.3)$$

Fixed effects models such as those specified in 2.1 to 2.3 are widely used within the econometric literature. Yet as Bell and Jones (2014) among others contend, the use of school fixed effects only considers the variance within schools over time and risks losing vital information on the variation between schools. Instead, random effect (multilevel) models could be applied, which would allow for unobserved time-invariant school-level variables to be modelled. However, this comes at the cost of making an additional assumption, the random effects assumption, which states that this unobserved heterogeneity is uncorrelated with the observed time-varying covariates *x.* This may be too strong an assumption given the nature of the data, a ten-year strongly balanced panel of repeated observations per school (Clarke et al, 2013). Nonetheless, it is possible to construct random effects models that do not make the random effects assumption. I present the results from such a model at Appendix 2. Whilst suggesting a similar overall effect of the programme to the fixed effects models, the results also reveal some variation between schools.

The credibility of the results presented depends on the selection of schools into 'treatment' and 'comparison'. This was by no means random. As described in Section 2, the order in which local authorities were selected to participate in the project was on the basis of lower attainment and higher levels of disadvantage. Authorities in later waves, those most at risk from the decision to halt the programme, were therefore by definition less disadvantaged and higher attaining than those which participated in earlier waves. Furthermore, we might have expected low attaining authorities (or low attaining schools for that matter) to improve simply as a result of reversion to the mean.

A check that the 'treatment' and 'comparison' groups were balanced in terms of observable characteristics at the start of the analysis window was performed by propensity score matching using the psmatch2 package in Stata (Leuven & Sianesi, 2003). The results of rebalancing are shown in Table 4. Variables used include the two end of Key Stage 4 outcome measures for 2003/04 and a number of measures of pupil characteristics sourced from the January 2004 School Census. The attainment measures relate to pupils in the final year of compulsory schooling (year 11) whilst the pupil characteristics measures additionally include years 7 to 10, who were assessed at the end of Key Stage 4 between 2004/05 and 2007/08 inclusive.

Table 4: Pupil Characteristics and Key Stage 4 attainment, 2003/04

		Actual	Rebalanced
% pupils in years 7 to 11 eligible for free school meals	Comparison	20%	25%
	Treatment	26%	26%
% pupils in years 7 to 11 with SEN met by School Action + or a statement	Comparison	8%	8%
	Treatment	9%	9%
% pupils in years 7 to 11 whose first language is not English	Comparison	12%	14%
	Treatment	15%	15%
% pupils in years 7 to 11 of White British or Irish ethnic background	Comparison	81%	77%
	Treatment	77%	77%
Mean Key Stage 2 fine grade of pupils in years 7 to 11	Comparison	4.43	4.34
	Treatment	4.32	4.32
% year 11 pupils achieving 5 or more A*-C grades (or equivalent) including GCSE English and mathematics	Comparison	36%	30%
	Treatment	29%	29%
Mean GCSE grade of year 11 pupils	Comparison	3.95	3.70
	Treatment	3.63	3.63

Selection issues may have affected the very final set of projects to reach 'financial close' and therefore to continue to the rebuilding stage. Some local authorities may have been better organised than others in this respect. However, it is unclear whether better project management would have any direct effect on pupil attainment. In any event, many of the final projects to be approved were still in progress at the end of the 2012/13 academic year and so are not included among the 'treatment' group in the analysis presented here.

Given the non-experimental nature of this evaluation, several identifying assumptions are made. Firstly, the stable unit treatment value assumption (or SUTVA; Rubin, 1980) which states that only the 'treated' benefit from the treatment. Conceivably, 'comparison' schools may have had to find other ways to improve in order to compete locally with schools rebuilt under BSF. A basic test of this assumption is presented in Section 5. Secondly, the 'difference-in-differences' specification assumes that the effects of unobserved variables (e.g. levels of funding) are time invariant. Thirdly, the common trends assumption (Blundell et al, 2004) is made, namely that patterns in attainment among 'treatment' and 'comparison' schools would have been the same in the absence of BSF. This is also tested to some extent in Section 5. Finally, the conditional independence assumption (Rubin, 1977) is made in the 'treatment vs comparison' specification. In the context of this research, this means assuming that, conditional on the set of covariates *x*, outcomes y are independent. In other words, there are 'comparison' schools similar to 'treatment' schools in terms of *x.* The 'rebalanced' set of 'comparison' schools attempts to take this a stage further by assuming that outcomes y are independent of observed covariates *x* and the propensity to be treated. The limitation of rebalancing is that the 'treatment' and 'comparison' groups may still differ on unobserved covariates which are correlated with the outcome.

5. Results

Table 5 presents summary outcome data for the full set of 'treatment' schools, those that were rebuilt or remodelled under BSF, and the 'comparison' group of schools whose projects were stopped. Charts 1 and 2 show some of this information visually. 'Treatment' schools include those schools whose projects were still in construction at the end of the 2012/13

year. The 'Policy On' column consists of schools whose BSF projects had been completed, the first of which was the West London Academy in 2005/06.

Table 5: Summary Key Stage 4 Attainment 2003/04 to 2012/13

KS4 Year	Number of pupils			% 5A*-C inc. English & maths			Mean GCSE Grad		
	Policy On	Treatment	Comparison	Policy On	Treatment	Comparison	Policy On	Treatment	Com
2003/04	0	114522	88019		29%	36%		3.64	
2004/05	0	112938	87221		31%	38%		3.75	
2005/06	111	114852	88672	27%	32%	40%	3.56	3.82	
2006/07	670	115777	89202	19%	35%	42%	3.37	3.93	
2007/08	1398	114363	88537	29%	38%	44%	3.49	4.05	
2008/09	5309	108982	85871	37%	41%	48%	3.92	4.15	
2009/10	15873	107527	85396	44%	47%	53%	4.14	4.28	
2010/11	27227	103956	83269	51%	51%	56%	4.38	4.34	
2011/12	44833	101539	82320	54%	53%	58%	4.42	4.39	
2012/13	69656	103387	83630	55%	55%	59%	4.40	4.41	

Chart 1: Percentage of pupils achieving 5 or more A*-C grades including English and mathematics, treatment and comparison schools

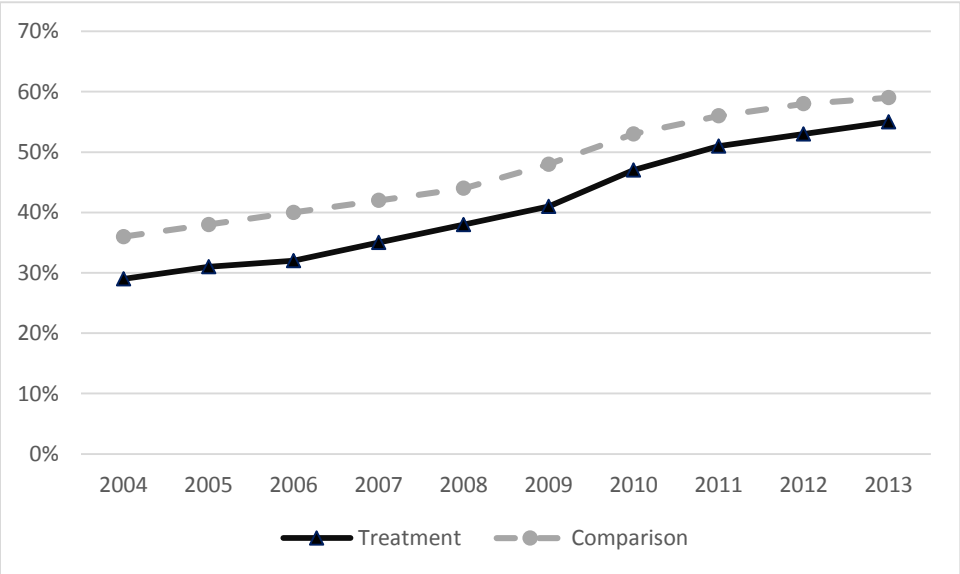
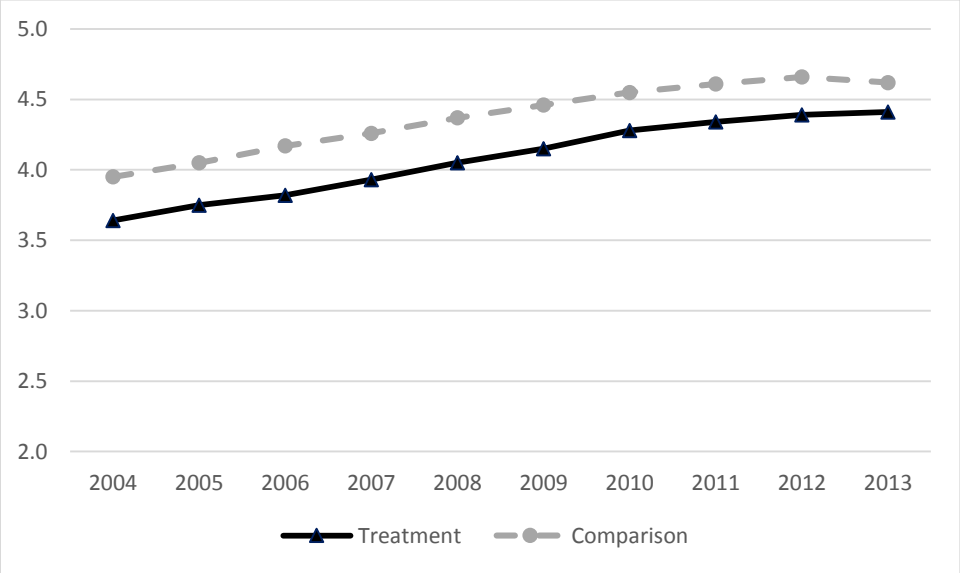


Chart 2: Mean GCSE grade, treatment and comparison schools



Prior to the opening of the first BSF school in 2005/06, average levels of attainment were higher among comparison schools compared to 'treatment' schools. Nonetheless, attainment improved at broadly the same rate, which lends some support to the common trends assumption. Levels of attainment in 'comparison' schools were still higher by the end of 2012/13 although gaps had narrowed slightly. Among the 'treatment' group, the percentage of pupils achieving 5 or more A*-C grades almost doubled between 2003/04 and 2012/13 and the mean GCSE grade improved by 0.8 grades.

In Table 6, 'treatment' schools have been split into five 'cohorts' based upon calendar year of opening. This suggests that attainment has improved at broadly similar rates in all types of 'treatment' school and comparison schools. As noted above, attainment tended to be lower among 'treatment' schools than comparison schools in 2003/04 (particularly among treatment schools which opened in 2011 or 2012) although average attainment in comparison schools was still lower than the national average for state-funded maintained schools.

Table 6: Mean GCSE grade by KS4 year and BSF cohort, treatment and comparison schools

KS4	BSF Cohort						
Year	>2013	2013	2012	2011	<=2010	Comparison	National
2003/04	3.85	3.56	3.59	3.66	3.66	3.95	4.18
2004/05	3.95	3.66	3.70	3.78	3.76	4.05	4.28
2005/06	4.01	3.73	3.77	3.83	3.83	4.17	4.35
2006/07	4.09	3.86	3.88	3.95	3.95	4.26	4.44
2007/08	4.23	3.95	4.00	4.06	4.09	4.37	4.56
2008/09	4.32	4.09	4.09	4.16	4.19	4.46	4.65
2009/10	4.44	4.21	4.21	4.32	4.31	4.55	4.74
2010/11	4.55	4.28	4.25	4.34	4.39	4.61	4.79
2011/12	4.57	4.33	4.29	4.38	4.44	4.66	4.81
2012/13	4.59	4.32	4.27	4.46	4.48	4.62	4.80

Excluding schools without 10 consecutive years of data from 2003/4, Tables 7a and 7b compare differences between treatment and comparison schools for both attainment measures, the proportion of pupils achieving 5 or more A*-C grades including English and mathematics (AC5EM) and mean GCSE grade. These parameters, estimated for each year separately, are null models. They show that treatment schools, save for those which were due to open in 2014 or later, tend to be lower attaining than comparison schools, and this was particularly so at the start of the window of analysis. The grey cells indicate years in which new or remodelled buildings were in use. A small number of the '<=2010' group opened in 2009 or earlier. Standard errors shown are clustered within schools. Charts 3 and 4 show the trend data visually for the 2012, 2011 and '<=2010' cohorts, with dashed lines representing years in which rebuilt schools were operational.

Table 7a: AC5EM mean differences by KS4 year and BSF cohort

KS4 Year		BSF Cohort				
		>2013	2013	2012	2011	<=2010
2003/04	Estimate	-0.022	-0.085	-0.079	-0.059	-0.057
	SE	0.026	0.016	0.015	0.019	0.016
2004/05	Estimate	-0.034	-0.082	-0.085	-0.064	-0.062
	SE	0.024	0.017	0.014	0.020	0.015
2005/06	Estimate	-0.033	-0.102	-0.092	-0.072	-0.067
	SE	0.025	0.016	0.013	0.019	0.016
2006/07	Estimate	-0.034	-0.097	-0.087	-0.060	-0.062
	SE	0.025	0.017	0.014	0.022	0.016
2007/08	Estimate	-0.030	-0.092	-0.082	-0.067	-0.050
	SE	0.023	0.018	0.014	0.020	0.016
2008/09	Estimate	-0.034	-0.086	-0.085	-0.061	-0.055
	SE	0.022	0.017	0.012	0.018	0.016
2009/10	Estimate	-0.027	-0.079	-0.077	-0.051	-0.052
	SE	0.023	0.017	0.014	0.019	0.016
2010/11	Estimate	-0.011	-0.068	-0.078	-0.051	-0.050
	SE	0.024	0.015	0.013	0.019	0.014
2011/12	Estimate	-0.013	-0.068	-0.075	-0.038	-0.039
	SE	0.022	0.014	0.012	0.017	0.014
2012/13	Estimate	0.000	-0.050	-0.059	-0.012	-0.035
	SE	0.020	0.014	0.012	0.017	0.015

Chart 3: AC5EM mean differences by KS4 year and BSF cohort

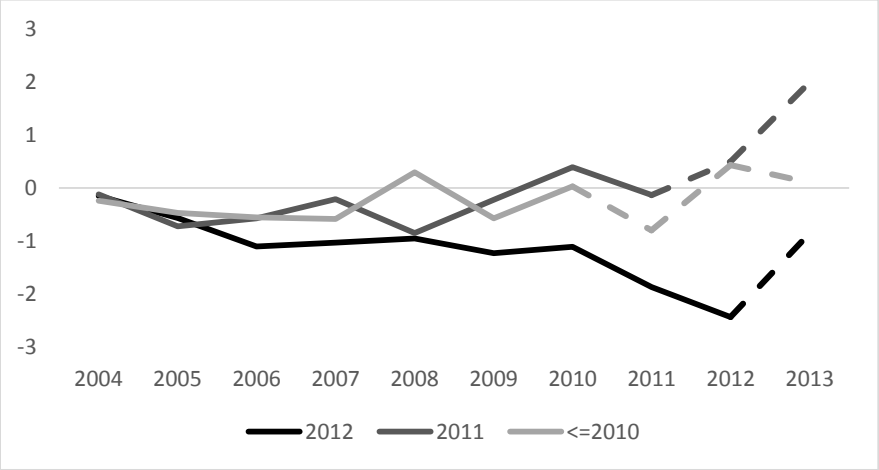
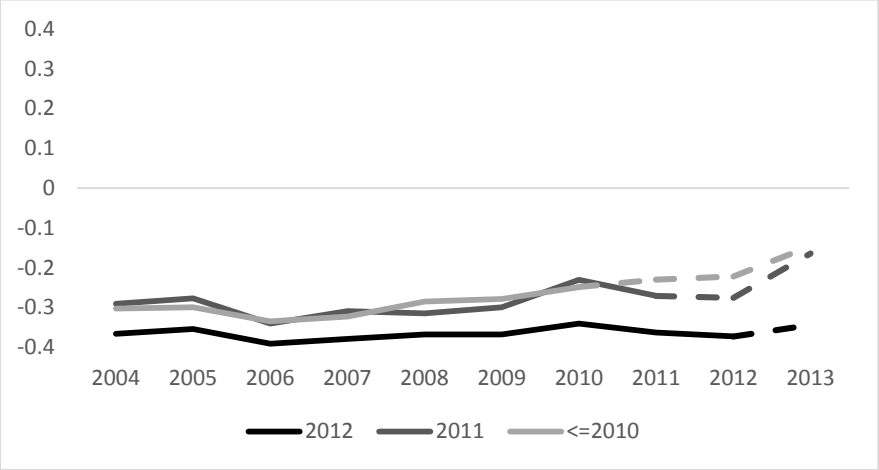


Table 7b: Mean GCSE Grade Differences by KS4 year and BSF cohort

KS4 Year		BSF Cohort >2013	2013	2012	2011	<=2010
2003/04	Estimate	-0.106	-0.397	-0.366	-0.291	-0.303
	SE	0.107	0.066	0.062	0.076	0.070
2004/05	Estimate	-0.110	-0.395	-0.354	-0.277	-0.300
	SE	0.102	0.064	0.058	0.077	0.070
2005/06	Estimate	-0.156	-0.439	-0.391	-0.341	-0.335
	SE	0.103	0.062	0.056	0.076	0.068
2006/07	Estimate	-0.177	-0.409	-0.379	-0.309	-0.323
	SE	0.100	0.059	0.056	0.071	0.066
2007/08	Estimate	-0.140	-0.422	-0.368	-0.315	-0.285
	SE	0.092	0.062	0.056	0.070	0.065
2008/09	Estimate	-0.142	-0.371	-0.368	-0.300	-0.279
	SE	0.088	0.062	0.053	0.071	0.064
2009/10	Estimate	-0.112	-0.337	-0.341	-0.231	-0.249
	SE	0.083	0.059	0.052	0.071	0.063
2010/11	Estimate	-0.069	-0.332	-0.363	-0.271	-0.230
	SE	0.084	0.063	0.050	0.073	0.062
2011/12	Estimate	-0.090	-0.327	-0.373	-0.276	-0.222
	SE	0.075	0.058	0.054	0.069	0.059
2012/13	Estimate	-0.028	-0.295	-0.344	-0.164	-0.146
	SE	0.078	0.061	0.052	0.075	0.060

Chart 4: Mean GCSE Grade Differences by KS4 year and BSF cohort



Differences in school and cohort characteristics can be controlled out (Tables 8a and 8b). Particularly for mean GCSE grade, these differences then tended to widen in the years prior to BSF projects being delivered. This tends to confirm the finding of the James Review that planning for BSF engendered a dip in school performance:

> [T]his level of input from the senior management team meant that attainment of pupils in their schools sometimes fell during and directly after the process. This is reflected in the views of many of the head teachers we spoke to during the course of this Review who felt the process was disruptive and used up far more of their time than was appropriate (James 2011, 2.35).

By the end of 2012/13, attainment on both outcome measures in treatment schools was not significantly different from that in comparison schools, except for mean GCSE grade among the 2012 cohort of treatment schools, which was significantly lower.

For the '<=2010' cohort of treatment schools, however, a non-significant positive difference is observed in 2012/13, which may suggest that any impact of a new school building on attainment is delayed. This may not be unreasonable as the first cohort of pupils to be examined at GCSE following completion of a school building project will have spent at most one year making use of the new facilities. By contrast, the 2012/13 cohort of pupils examined in the '<=2010' cohort of schools will have spent at least three years making use of them. Charts 5 and 6 show the trend data visually for the 2012, 2011 and '<=2010' cohorts, with dashed lines representing years in which rebuilt schools were operational.

Table 8a: AC5EM differences, controlling for school and cohort characteristics

KS4 Year		BSF Cohort >2013	2013	2012	2011	<=2010
2003/04	Estimate	0.094	-0.799	-0.146	-0.120	-0.239
	SE	1.226	0.791	0.718	0.855	0.751
2004/05	Estimate	-1.029	-0.945	-0.561	-0.718	-0.470
	SE	0.959	0.765	0.728	0.796	0.734
2005/06	Estimate	-1.247	-2.086	-1.101	-0.567	-0.549
	SE	1.016	0.731	0.737	1.009	0.733
2006/07	Estimate	-0.951	-1.640	-1.025	-0.208	-0.580
	SE	0.883	0.762	0.753	1.010	0.733
2007/08	Estimate	0.019	-1.261	-0.952	-0.846	0.296
	SE	0.906	0.877	0.814	0.950	0.808
2008/09	Estimate	-0.073	-0.500	-1.226	-0.219	-0.568
	SE	0.993	0.859	0.661	0.880	0.976
2009/10	Estimate	-0.387	-0.336	-1.105	0.397	0.030
	SE	1.030	0.840	0.782	0.926	0.991
2010/11	Estimate	0.209	0.174	-1.864	-0.131	-0.798
	SE	0.992	0.798	0.726	0.855	0.853
2011/12	Estimate	0.008	-0.900	-2.428	0.501	0.435
	SE	1.122	0.926	0.759	1.103	0.832
2012/13	Estimate	0.765	1.042	-0.850	1.991	0.101
	SE	1.144	0.930	0.789	1.192	0.877

Chart 5: AC5EM differences, controlling for school and cohort characteristics

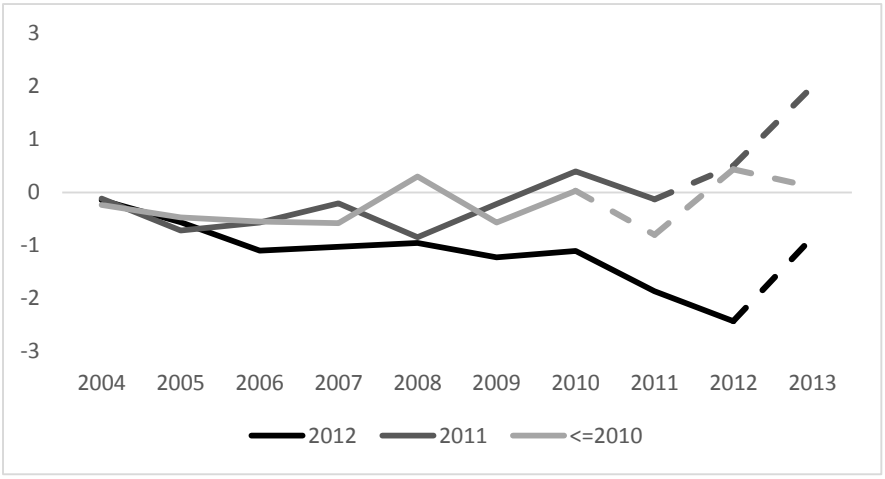
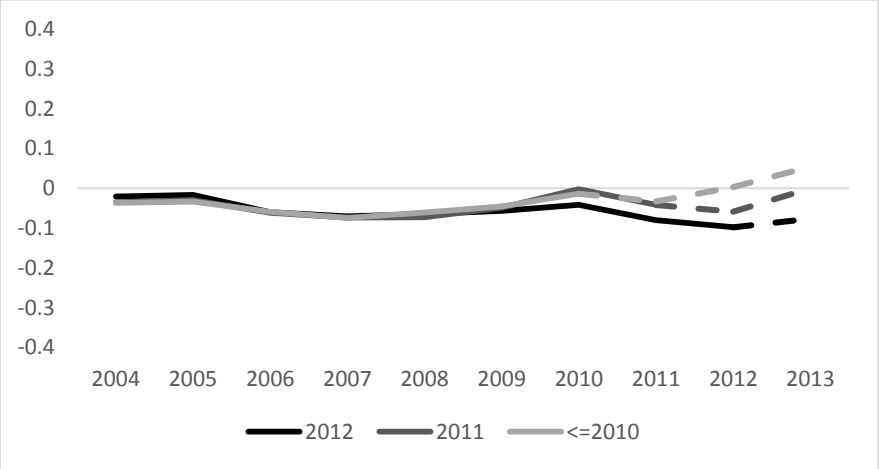


Table 8b: Mean Grade Differences, controlling for school and cohort factors

KS4 Year		BSF Cohort >2013	2013	2012	2011	<=2010
2003/04	Estimate	-0.010	-0.061	-0.021	-0.035	-0.037
	SE	0.055	0.037	0.036	0.041	0.040
2004/05	Estimate	-0.005	-0.083	-0.017	-0.030	-0.034
	SE	0.042	0.040	0.034	0.038	0.038
2005/06	Estimate	-0.074	-0.109	-0.060	-0.062	-0.059
	SE	0.047	0.035	0.036	0.049	0.037
2006/07	Estimate	-0.078	-0.083	-0.071	-0.074	-0.075
	SE	0.046	0.036	0.033	0.044	0.032
2007/08	Estimate	-0.014	-0.101	-0.065	-0.073	-0.061
	SE	0.037	0.035	0.032	0.042	0.028
2008/09	Estimate	-0.001	-0.027	-0.057	-0.049	-0.046
	SE	0.039	0.033	0.030	0.038	0.034
2009/10	Estimate	-0.011	-0.006	-0.042	-0.002	-0.014
	SE	0.032	0.033	0.030	0.034	0.037
2010/11	Estimate	-0.006	-0.012	-0.080	-0.042	-0.033
	SE	0.031	0.037	0.026	0.032	0.030
2011/12	Estimate	-0.039	-0.036	-0.098	-0.059	0.003
	SE	0.032	0.032	0.030	0.034	0.029
2012/13	Estimate	0.001	0.009	-0.077	0.000	0.054
	SE	0.033	0.033	0.029	0.037	0.029

Chart 6: Mean Grade Differences, controlling for school and cohort factors



The viability of the SUTVA assumption was tested by comparing the attainment of 'comparison' schools that were spatially proximate (within 5km) to 'treatment' schools to attainment at 'comparison' schools that were less proximate (at least 10km from a 'treatment' school). Any improvement over time in the attainment of the former group relative to the latter group may indicate a violation of SUTVA arising from proximate schools finding other means to compete with rebuilt schools. The basic treatment vs comparison of model 1.1 is used, with a Boolean proximity flag acting as the treatment effect. The results in Table

9 suggest that spatial proximity to BSF schools has, if anything, a negative impact on pupil attainment.

Table 9: Differences in attainment between 'comparison' schools proximate to BSF schools and those less proximate, controlling for school and cohort factors

	Mean GCSE grade		% achieving 5 or more A*-C (inc Eng & mat)	
	Estimate	SE	Estimate	SE
2003/04	-0.01	0.05	-0.4%	1.1%
2004/05	-0.01	0.05	-0.7%	1.1%
2005/06	-0.06	0.05	-0.2%	1.1%
2006/07	-0.03	0.06	-0.9%	1.3%
2007/08	-0.04	0.05	-0.5%	1.4%
2008/09	-0.05	0.05	-2.1%	1.1%
2009/10	-0.04	0.04	0.6%	1.2%
2010/11	-0.04	0.04	-0.7%	1.2%
2011/12	-0.06	0.04	-1.6%	1.3%
2012/13	-0.09	0.04	-1.3%	1.3%

We now turn to the difference-in-difference models specified in equations 2.1 to 2.3. These are fitted both with and without controls. Overall (2.1), these show no significant impact of new school buildings on attainment. However, there is a significant impact on AC5EM (Table 9a, equation 2.3) among the 2011 cohort of BSF schools, particularly in the 2012/13 school year. Additionally, there is a significant impact on mean GCSE grade (Table 9b, equation 2.3) among the '<=2010' cohort in 2012/13. Almost identical results are obtained by rebalancing the sample using propensity score matching (Appendix 2). These estimates are slightly larger than those estimated using the 'treatment vs comparison' specification of equation 1.1 reported in Tables 8a and 8b.

Table 9a: Difference-in-Differences Estimates (school level data), AC5EM

			Without Controls		With Controls	
Equation	KS4 Year	BSF Cohort	Estimate	SE	Estimate	SE
2.1	All	All	0.90	0.51	0.70	0.46
2.2	All	2012	1.28	0.80	0.85	0.73
	All	2011	2.56	1.04	2.07	0.99
	All	<=2010	0.11	0.68	0.09	0.60
2.3	2010/11	<=2010	-0.43	0.74	-0.35	0.67
	2011/12	2011	1.72	1.09	1.37	1.03

	2011/12	<=2010	0.65	0.92	0.74	0.86
	2012/13	2012	1.52	0.82	1.00	0.75
	2012/13	2011	3.65	1.19	2.94	1.14
	2012/13	<=2010	0.72	0.94	0.33	0.85

Table 9b: Difference-in-Differences Estimates (school level data), Mean GCSE Grade

			Without Controls		With Controls	
	KS4 Year	BSF Cohort	Estimate	SE	Estimate	SE
2.1	All	All	0.03	0.02	0.02	0.02
2.2	All	2012	-0.02	0.02	-0.04	0.02
	All	2011	0.04	0.03	0.02	0.03
	All	<=2010	0.03	0.03	0.04	0.02
2.3	2010/11	<=2010	0.02	0.03	0.02	0.03
	2011/12	2011	0.00	0.03	-0.01	0.03
	2011/12	<=2010	0.03	0.03	0.04	0.03
	2012/13	2012	0.00	0.02	-0.02	0.02
	2012/13	2011	0.08	0.03	0.06	0.03
	2012/13	<=2010	0.10	0.03	0.09	0.03

Equations 3.1 to 3.3 specify models analogous to 2.1 to 2.3 based on pupil-level data. Results are shown in Tables 10a and 10b. The probability of achieving 5 or more A*-C grades including English and mathematics is modelled using logistic regression. Results broadly similar to the school-level analysis shown in Table 9b are produced. For mean GCSE grade, a small overall effect for the '<=2010' cohort is found (Table 10b). However, it is extremely small and may simply be the result of reversion to the mean of previously low attaining schools. In 2012/13, the national standard deviation in pupil-level mean GCSE scores was 1.5 (0.6 at school-level). The difference of 0.05 from equation 3.2 shown in Table 10b corresponds to an effect size of around 0.03 with a 95% confidence interval of 0.02. There is a significant effect of achieving 5 or more A*-C grades including English and mathematics among treated schools, largely driven by the 2011 cohort of schools (Table 10a).

Table 10a: Difference-in-Differences Estimates (pupil level data), odds-ratio of achieving AC5EM

			Without Controls		With Controls	
Equation	KS4 Year	BSF Cohort	Estimate	SE	Estimate	SE
3.1	All	All	0.041	0.022	0.083	0.039
3.2	All	2012	0.045	0.034	0.075	0.058
	All	2011	0.108	0.045	0.160	0.080
	All	<=2010	0.011	0.031	0.052	0.052
3.3	2010/11	<=2010	-0.013	0.033	-0.004	0.058
	2011/12	2011	0.074	0.047	0.115	0.085
	2011/12	<=2010	0.034	0.041	0.116	0.075
	2012/13	2012	0.054	0.035	0.082	0.060

	2012/13	2011	0.153	0.052	0.217	0.092
	2012/13	<=2010	0.031	0.041	0.051	0.069

Table 10b: Difference-in-Differences Estimates (pupil level data), Mean GCSE Grade

			Without Controls		With Controls	
	KS4 Year	BSF Cohort	Estimate	SE	Estimate	SE
3.1	All	All	0.03	0.02	0.03	0.02
3.2	All	2012	-0.02	0.02	-0.03	0.02
	All	2011	0.04	0.03	0.02	0.03
	All	<=2010	0.04	0.03	0.05	0.02
3.3	2010/11	<=2010	0.02	0.03	0.03	0.03
	2011/12	2011	0.00	0.03	0.00	0.03
	2011/12	<=2010	0.03	0.03	0.05	0.03
	2012/13	2012	0.00	0.02	-0.01	0.02
	2012/13	2011	0.08	0.03	0.05	0.03
	2012/13	<=2010	0.10	0.03	0.09	0.03

Finally, I briefly analyse changes in school composition as a result of BSF by running a difference-in-differences model without controls (equation 3.1) with the percentage of year 7 pupils eligible for free school meals (FSM) as the dependent variable. Table 11 indicates a small reduction in eligibility among the earliest BSF cohorts. Based on equation 3.3, it would appear that the proportion of disadvantaged pupils enrolling at 'treated' schools has been reducing over time.

Table 11: Difference-in-Differences Estimates (school level data), Percentage of year 7 pupils known to be eligible for free school meals

	KS4 Year	BSF Cohort	Estimate	SE
3.1	All	All	0.96	0.31
3.2	All	2012	0.27	0.60
	All	2011	-1.86	0.65
	All	<=2010	-1.32	0.46
3.3	2010/11	<=2010	-0.98	0.53
	2011/12	2011	-1.73	0.70
	2011/12	<=2010	-1.67	0.58
	2012/13	2012	0.34	0.61
	2012/13	2011	-1.87	0.78
	2012/13	<=2010	-1.81	0.61

6. Discussion

The James Review of Education Capital found 'very little evidence that a school building that goes beyond being fit-for-purpose has the potential to drive educational transformation. The generally held view was that the quality of teachers and leaders has a much greater impact on attainment than the environment' (James, 2011, 2.7). The analysis presented here based

on a number of model specifications tends to suggest that, on average, new school buildings have little, if any, effect on attainment at Key Stage 4, at least in the short run. It is possible that new school buildings have improved wider outcomes such as attendance, enjoyment of learning or the safety of children, some of which can be measured using administrative data. They may have also provided facilities that can be used more widely by local communities.

There is a suggestion of a delayed effect of new school buildings on attainment among some of the earliest BSF projects, which may well increase in magnitude over time. Nonetheless, the current sizes of these effects on attainment are small, particularly in comparison to low-cost teacher interventions such as giving feedback to pupils[vi]. Given the lower level of attainment among the earlier BSF projects, this finding may be no more than reversion to the mean.

Moreover, the proportion of pupils enrolled into year 7 from disadvantaged backgrounds, measured by eligibility for free school meals, tended to fall among 'treated' schools relative to 'comparison' schools. This does not necessarily mean that such pupils have been displaced by more advantaged pupils. It may be the case that previously unpopular, under-subscribed schools have been more successful in recruiting pupils. Nonetheless, BSF does appear to have had more impact on school composition than on attainment.

On the surface, the absence of an effect on attainment in this research does not accord with that conducted by Neilson and Zimmerman (2011) in the USA or Barrett et al (2014) in England. However, neither of these studies examined impacts on secondary (high) school students. The former analysed the impact of new elementary and middle school buildings in a poor, urban area (New Haven, Connecticut) where 80% of pupils were eligible for a free lunch and 90% were from Black or Hispanic ethnic backgrounds. Although local authority areas with higher levels of deprivation were prioritised under BSF, their pupil populations were much more diverse (see Table 4). Secondly, the schools rebuilt in New Haven had been subject to a common set of design requirements (e.g. heating and air conditioning, improved classroom technology). By contrast, each local authority defined its own requirements for BSF.

Barrett et al (2014) find large impacts of classroom design on pupil progress in primary schools. This research does not necessarily contradict this finding. The schools rebuilt under BSF may not have been any better designed than the schools whose BSF projects were cancelled. Secondly, Barrett et al analyse classrooms rather than schools and uncover substantial within-school variation, i.e. a school may have well-designed and poorly-designed classrooms. Finally, the progress measures on which their findings are predicated appear to be based on teacher assessments. Some of the between-classroom variations in progress they report may be the result of variation in assessment between teachers (inter-rater reliability).

As Mansell (2011) argues, there was an inherent tension between localism and centralism at the heart of new labour educational policy. The debate on whether schools operate as a market or as a closed system persists to this day (Allen & Burgess, 2011). On the one hand, City Academies (City was later dropped from the name) and BSF avowed to involve local communities and businesses in the commissioning of schools. On the other, the Department for Education and Skills (DfES) went beyond maintaining and developing the national curriculum inherited from the previous administration by directing *how* it should be taught through the Primary and Secondary National Strategies (Ofsted, 2008). A Standards and Effectiveness Unit (SEU) was established to monitor its delivery, including the statutory setting of school, local authority and national targets. Schools whose performance fell below "floor standards" were "named and shamed" (Mansell, 2011). The Office for Standards in

Education inspected all schools using the same framework. And although BSF may have given some schools the opportunity to expand their curriculum by including facilities for vocational subjects such as Motor Vehicle Studies or Hairdressing, the preponderant qualifications taught in schools, GCSEs, remained largely unchanged. In retrospect then, it is difficult to imagine how much local transformation could take place in this era of unprecedented 'top-down' management of schools by central government. It is perhaps only to be expected that the estimated effects of BSF presented in this paper are marginal at best.

The small effect noted for 2012/13 among the earliest cohort of BSF schools may be worth revisiting at a later stage. It is conceivable that the effects of a new school are lagged rather than instant. The first cohort of pupils to reach the end of Key Stage 4 in a rebuilt school will have spent just a matter of months making use of the new facilities having spent up to four years being taught in the predecessor school and possibly experiencing disruption as a result of building works in their penultimate year of compulsory schooling. This hypothesis can be explored further by repeating the analysis in the future. The impact of new schools may have been attenuated by the relatively permissive way BSF was implemented, which allowed some poorly designed schools to be built (James, 2011). Moreover, if there are small annual effects on achievement resulting from rebuilding new schools, these effects may well persist for a number of years (Martorell et al, 2015).

The dataset on schools rebuilt under BSF may illuminate the debate on why the first wave of Academies was apparently more effective than later waves (Machin & Vernoit, 2011). Almost all of those in the first wave were rebuilt or extensively remodelled but those that followed were much less likely to be so. By comparing rebuilt early Academies to similar schools which were rebuilt but which did not become Academies (initially, at least), the effect of the change of governance may be isolated from the effect of the new school buildings.

Further data on the condition of the school estate in England has recently become available as a result of the Property Data Survey Programme (PDSP) introduced following the James Review (Education Funding Agency, 2012). Through a comprehensive set of surveys completed by August 2014, the programme aims to produce up-to-date data on the condition of all 23 thousand schools in England. By exploiting variations in the quality of conditions between schools, this new data source may yield new insights into the effects of school buildings on attainment.

Appendix 1: Results of 'within-between' random effects models

In order to examine the variation between schools in the effect of Building Schools for the Future, I construct a random effects model in which the random effects assumption is not made. This is based on the the 'within-between' formulation proposed by Bell and Jones (2014) shown in 4.1.

$$y_{st} = \delta policyon_{st} + \tau\overline{policyon}_s + \beta(x_{st} - \bar{x}_s) + \alpha\bar{x}_s + \mu_s + e_{st} \quad (4.1)$$

In this specification, *s* represents a school and t an annual observation over the period 2003/04 to 2012/13. Each time-varying school-level covariate $x_{st}$is centred on its mean $\bar{x}_s$ over the time period observed. The means of all time-varying school-level covariates are also included in the model. Similarly, parameters are estimated for the time-varying *policyon* variable along with its time-invariant mean $\overline{policyon}$ which represents the average effect on attainment over the time period observed of being a 'treated' school. μ_s represents the random time-invariant school effect and e_{st} the error term for each school in each year. These models were fitted using the xtmixed command in Stata (StataCorp, 2012) in preference to alternatives since it permits the use frequency weights, in this case the number of pupils in each school in each year.

Such models yield two sets of estimates. Firstly, 'within' estimates for the time-varying covariates, which are consistent with those from a fixed effects model (e.g. 2.1). Secondly, 'between' estimates for the school-level means, which represent the effect of average levels of *x* and *policyon* over the period observed. However, compared to a 'standard' random effects model in which the random effects assumption is made, the level 2 (school-level) effects (and therefore their variance) are affected (Snijders & Berkhof, 2007) since the unobserved heterogeneity at school level is confounded with the non-linear effects of the school means (Clarke et al, 2013; Castellano et al, 2014). This would be a particular concern if interpretation of the level 2 effects was of substantive interest (Fielding, 2004).

A 'within-between' random effects model (equation 4.1) was fitted for the mean GCSE grade outcome using the school level dataset. As noted above and shown in Table 12, it yields identical parameter estimates to its fixed effects analogue (equation 2.1). It additionally partitions the residual variance into within and between school components. This shows that approximately half the residual variance is between schools. This variance is not considered at all in the fixed effects model. The 'within-between' model can be extended further to examine differential effects of the time-varying *policyon* variable or any of the covariates.

Table 12: 'Within-between' estimates and variance components, mean GCSE grade

	Variance Components		Random Coefficient	
Parameter	Estimate	SE	Estimate	SE
Fixed effect: policyon (time-varying)	0.02	0.02	0.02	0.02
Fixed effect: policyon school mean (time-	-0.07	0.06	-0.04	0.06
Within-school variance	0.038	0.001	0.034	0.001
Between-school variance (random intercepts)	0.050	0.002	0.049	0.002
Between-school variance (random policyon			0.074	0.007
Between-school variance (covariance)			-0.009	0.003

Fitting a random coefficient for *policyon* reveals significant variation between schools (Table 12). This tends to suggest that new school buildings may have positive effects on pupil attainment under certain conditions. To explore this further, I introduce an additional time-invariant fixed effect, the percentage of pupils achieving 5 or more A*-C grades at the start of the observation window (2002/03) along with its interaction with *policyon*. Results are shown in Table 13. The interaction term achieves borderline statistical significance at the 95% level but the parameter estimate is positive. This suggests that it was schools with higher levels of attainment at the start of the programme that tended to achieve greater benefits from new school buildings.

Table 13: Parameter estimates from within-between model

Parameter Type	Parameter	Coef.	Robust std error
Year dummies	2004	-0.03	0.01
(2003 is reference category)	2005	0.04	0.01
	2006	0.12	0.01
	2007	0.24	0.01
	2008	0.30	0.01
	2009	0.37	0.01
	2010	0.39	0.02
	2011	0.40	0.02
	2012	0.36	0.02
Time varying effects	Policyon	0.019	0.016
centred around school means	Cohort mean KS2 grade	6.142	1.079
	Cohort st dev KS2 fine grade	4.171	1.055
	%FSM	0.171	0.974
	%White	0.034	0.064
	%EAL	0.258	0.061
	Cohort mean KS2 grade squared	-0.505	0.110
	Interaction: %FSM and %White	-1.173	0.186
	Interaction: KS2 mean and %FSM	0.074	0.226
	Interaction: KS2 mean and KS2 st dev	-0.919	0.238
School means	Policyon	-0.222	0.087
of time varying effects	Cohort mean KS2 grade	-9.973	1.950
	Cohort st dev KS2 fine grade	-13.264	1.933
	%FSM	4.691	2.047
	%White	-0.117	0.101
	%EAL	0.293	0.111
	Cohort mean KS2 grade squared	1.098	0.190
	Interaction: %FSM and %White	-0.978	0.241
	Interaction: KS2 mean and %FSM	-1.060	0.480
	Interaction: KS2 mean and KS2 st dev	2.945	0.429
Time invariant effects	Northern regions	-0.037	0.027
	Midlands regions	-0.011	0.027
	Southwest regions	0.011	0.035
	Eastern regions	-0.022	0.031
	London	0.032	0.033
	Girls only school	0.210	0.181
	Boys only school	0.009	0.126
	%AC5EM 2003	1.443	0.153
	Constant	3.505	0.045
Time invariant*policy on interaction	policyon and %AC5EM 2003	0.526	0.273

Appendix 2: Results of Difference-in-Differences Models Controlling for Propensity Score Weights

Table 14a: Difference-in-Differences Estimates (school level data), AC5EM

Equation	KS4 Year	BSF Cohort	Without Controls		With Controls	
			Estimate	SE	Estimate	SE
2.1	All	All	0.14	0.52	0.38	0.48
2.2	All	2012	0.65	0.81	0.71	0.75
	All	2011	1.82	1.04	1.77	1.00
	All	<=2010	-0.71	0.68	-0.31	0.62
2.3	2010/11	<=2010	-1.15	0.75	-0.74	0.69
	2011/12	2011	0.93	1.10	0.97	1.04
	2011/12	<=2010	-0.40	0.93	0.15	0.88
	2012/13	2012	0.83	0.83	0.88	0.77
	2012/13	2011	2.85	1.20	2.74	1.15
	2012/13	<=2010	-0.28	0.95	0.06	0.88

Table 14b: Difference-in-Differences Estimates (school level data), Mean GCSE Grade

	KS4 Year	BSF Cohort	Without Controls		With Controls	
			Estimate	SE	Estimate	SE
2.1	All	All	0.00	0.02	0.01	0.02
2.2	All	2012	-0.05	0.02	-0.05	0.02
	All	2011	0.01	0.03	0.01	0.03
	All	<=2010	0.01	0.03	0.02	0.02
2.3	2010/11	<=2010	-0.01	0.03	0.01	0.03
	2011/12	2011	-0.02	0.03	-0.02	0.03
	2011/12	<=2010	0.00	0.03	0.03	0.03
	2012/13	2012	-0.03	0.03	-0.03	0.02
	2012/13	2011	0.05	0.03	0.05	0.03
	2012/13	<=2010	0.05	0.03	0.07	0.03

References

Allen, R. & Burgess, S. (2011) The future of competition and accountability in education, *CMPO Working Paper*, University of Bristol, Retrieved from http://www.bristol.ac.uk/cmpo/publications/other/competition.pdf

Barrett, P., Davies, F., Zhang, Y. & Barrett, L. (2014) The impact of classroom design on pupils' learning: Final results of a holistic mutli-level analysis, *Building and Environment,* 89 (2014) 118-133.

Bell, A. & Jones, K. (2014) Explaining Fixed Effects: Random Effects modelling of Time-Series Cross-Sectional and Panel Data, *Political Science Research and Methods*, 2014, 1-21.

Blundell, R., Costa Dias, M., Meghir, C. & van Reenen, J. (2004) Evaluating the employment impact of a mandatory job search program, *Journal of the European Economic Association,* 2 (4), 569 – 606.

Castellano, K., Rabe-Hesketh, S. & Skrondal, A. (2014) Composition, context and endogeneity in school and teacher comparisons, *Journal of Education and Behavioural Statistics,* 39 (5), 333-367

Clarke, P., Crawford, C., Steele, F. & Vignoles, A. (2013) Revisiting fixed- and random-effects models: some considerations for policy-relevant education research. *Education Economics*, 2013, 1-19.

Department for Education (2012) *The national pupil database user guide* (London, DfE). Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/261189/NPD_User_Guide.pdf

Edgerton, E., McKechnie, J. & McEwen, S. (2010) Building better schools: methodological concerns and the need for evidence-based research in E. Knapp, K. Noschis (Eds) *Architectural Quality in Planning and Design of Schools* (Lausanne, Imprimerie Chabloz).

Edgerton, E., McKechnie, J. & McEwen, S. (2011) Students' perceptions of their school environments and the relationship with educational outcomes, *Educational & Child Psychology*, 28 (1), 33-45.

Education Funding Agency (2012) *Property Survey Data Programme: Survey Manual Part 1* (Coventry, EFA). Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/276415/pdsp_survey_manual_part_1_a4_revision_20_v9_final_access.pdf

Fielding, A. (2004) The role of the Hausman test and whether higher level effects should be treated as random or fixed, *Multilevel modelling newsletter*, 16 (2), 3-9.

Foucault, M. (1975) *Discipline and Punish,* (New York, Random House).

Foley, B. & Goldstein, H. (2012) *Measuring success: League tables in the public sector*, (London, the British Academy).

Higgins, S., Hall, E., Wall, K., Woolner, P. & McCaughey, C. (2005) *The impact of school environments: A literature review* (London, Design Council).

Hobbs, G. & Vignoles, A. (2009) 'Is children's free school meal 'eligibility' a good proxy for family income?', *British Educational Research Journal,* 1469-3518.

House of Commons Treasury Committee (2012) *Private Finance Initiative*, Seventeenth Report of Session 2010-2012 (London, Stationery Office).

James, S. (2011) *Review of Education Capital* (London, DfE). Retrieved from https://www.education.gov.uk/consultations/downloadableDocs/James%20Reviewpdf.pdf

Jenkins, A., Levačić, R., & Vignoles, A. (2006) *Estimating the relationship between school resources and pupil attainment at GCSE* (London, DfES).

Leiringer, R. & Cardellino, P. (2011) Schools for the twenty-first century: school design and educational transformation *British Educational Research Journal*, 36(7), 915-934.

Leuven E. & Sianesi B. (2003) psmatch2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Retrieved from http://ideas.repec.org/c/boc/bocode/s432001.html

Machin, S. & Vernoit, J. (2011) Changing school autonomy: Academies and their introduction to England's education, *CEE discussion paper dp 123*, (London, CEE).

Mahony, P., Hextall, I. & M. Richardson (2011) 'Building Schools for the Future': reflections on a new social architecture, *Journal of Education Policy,* 26(3), 341-360

Mahony, P. & Hextall, I. (2012) 'Building Schools for the Future': 'transformation' for social justice or expensive blunder?, *British Educational Research Journal,* 39(5), 853-871.

Mansell, W. (2011) Improving exam results, but to what end? The limitations of New Labour's control mechanism for schools: assessment-based accountability, *Journal of Educational Administration and History*, 43(4), 291-308.

Martorell, F., Strange, K., McFarlin, I. (2015) Investing in Schools: Capital Spending, Facility Conditions, and Student Achievement, *NBER discussion paper no. 21515,* (Massachusetts, NBER)

National Audit Office (2009) *The Building schools for the future programme: Renewing the secondary school estate* (London, The Stationery Office).

Neilson, C. & Zimmerman, S. (2011) The effects of school construction on test scores, school enrolment and house prices, *IZA discussion paper dp6106,* (Bonn, IZA).

Ofsted (2008) *Evaluation of the Primary and Secondary National Strategies* (London, Ofsted).

PfS/4ps (2008) *An introduction to Building Schools for the Future* (London, 4ps and PfS).

PricewaterhouseCoopers (2010) *Evaluation of Building Schools for the Future – 3rd annual report* (London, DCSF).

Rubin, D.B. (1977) Assignment to Treatment Group on the Basis of a Covariate, *Journal of Educational Statistics*, 2, 1-26.

Rubin, D. B. (1980) Comment on: Randomization analysis of experimental data in the Fisher randomization test by D. Basu, *Journal of the American Statistical Association,* 75, 591-593.

Snijders, T. A. B. & Berkhof, J. (2007) Diagnostic checks for multilevel models, in J. de Leeuw & E. Meijer (Eds) *Handbook of Multilevel Analysis*, 139-73, (New York, Springer).

StataCorp (2012) Xtmixed: Multilevel mixed-effects linear regression (Texas, Stata Press).

West, D., Barham, E. & Hind, A. (2009) *Secondary School Admissions in England: Policy and Practice* (London, LSE).

Wolf, A. (2011) Review of Vocational Education- the Wolf report, (London, DfE).

[i] The final list of projects can be found at https://www.whatdotheyknow.com/request/55026/response/142942/attach/html/3/impact%20of%20building%20schools%20for%20the%20future%20announcement%20of%20monday%205%20july%202010.pdf.html (retrieved 22nd May 2014)

[ii] Each local authority selected up to 2 'sample' (or 'pilot') projects that were considered representative of the type of schools they planned to build. Construction on these projects began before non-sample projects.

[iii] First collected in 2001/2, School Census is an electronic census of all pupils enrolled at state-funded schools. Originally collected annually, it has been collected three times per year since 2006/7. The scope of the collection has changed over the year, with schools providing data on pupils, their characteristics (such as ethnicity and gender), home address, special educational needs, exclusions and attendance

[iv] Often referred to as 'league tables', performance data on secondary schools in England has been published annually since 1992/93. See http://www.education.gov.uk/schools/performance

[v] An example here would be the inclusion of international GCSEs (IGCSEs) in Performance Tables in 2009/10.

[vi] http://educationendowmentfoundation.org.uk/toolkit/feedback/