Department of Quantitative Social Science

Does school autonomy improve educational outcomes? Judging the performance of foundation secondary schools in England

Rebecca Allen

DoQSS Working Paper No. 1002
February 2010

Leading education and social research
Institute of Education
University of London

DISCLAIMER

Any opinions expressed here are those of the author(s) and not those of the Institute of Education. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

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

Department of Quantitative Social Science. Institute of Education, University of London. 20 Bedford way, London WC1H 0AL, UK.

Does school autonomy improve educational outcomes? Judging the performance of foundation secondary schools in England

Rebecca Allen[*†]

Abstract. Government and researchers use school performance measures such as contextual value-added to claim that giving schools autonomy from local authority control produces superior pupil performance in GCSE examinations. This paper explores the extent to which inferring causality between autonomy and pupil achievement is reasonable given that pupils are not randomly assigned to schools and schools do not randomly acquire autonomous status. Rich administrative data and the Longitudinal Survey of Young People in England are used to evaluate whether CVA-style inferences are confounded by pupil characteristics that explain both the chances of attending an autonomous school and academic achievement. The assignment of grant-maintained (and thus now foundation) status through a vote of parents is used to compare school that just did, and just did not, gain autonomy over a decade ago. These alternative estimation strategies suggest there is little evidence that foundation status casually yields superior school performance.

JEL classification: I21, I28.

Keywords: school autonomy, school effectiveness, foundation schools.

[*]Department of Quantitative Social Science, Institute of Education, University of London. 20 Bedford Way, London WC1H 0AL, UK. E-mail: (r.allen@ioe.ac.uk)

[†]**Acknowledgements**. Thanks are due to the Department of Children, Schools and Families for providing the data.

1. Introduction

Education policy of the two major political parties in England is currently predicated on assumption that giving schools greater autonomy from local authority control is a route to improving standards. This might be through more efficient decision-making and resource usage or because autonomy is a necessary precursor to market-like reforms whereby schools are somehow incentivised to compete for pupils. Autonomy from local government has always been a feature of English schooling through the presence of state-funded church schools from the 19th century onwards, but the past two decades have seen a rapid growth in autonomy through the grant-maintained schools policy, City Technology Colleges, Academies and new regulation to make it easier for schools to acquire foundation status. Today, 34 per cent of secondary school pupils are educated in an autonomous school.

Tables of school performance in England are used as the empirical evidence that autonomous schools achieve superior performance in GCSE exams taken by 16 year-olds (e.g. DfES, 2007). In these tables school performance is measured using contextual value-added (CVA) models where pupil achievement at GCSE is regressed on prior achievement and a set of pupil and school characteristics to control for context. Any unexplained variation is attributed to school quality. However, although policy makers and journalists tend to draw causal inferences from these tables, the methods used will only yield unbiased estimates of the effectiveness of school types if unmeasured social characteristics of the pupil intake are uncorrelated with type of school attended. Furthermore, to infer *causality* between autonomous status and academic achievement there should be no unmeasured area characteristics that determine both the chances of a school becoming autonomous and a school's ex-ante effectiveness.

There are different types of autonomous schools in the English education system today and this article chooses to focus on foundation schools. This large group of secondary schools is now long enough established to facilitate investigation of the long-run effects of school autonomy and it is

possible to do this without confounding other factors such as the religious ethos of the school. By contrast the Academy programme, whilst currently at the forefront of the policy debate, is too recent to evaluate stable state impacts, and in any case there is no straightforward route to identifying a comparison group of non-Academies (see Machin and Wilson, 2008, for a short-run impact evaluation of Academies).

There is no perfect approach to establishing causality where the assignment of autonomy to schools and the assignment of pupils to schools are both non-random. In this article a combination of non-experimental and quasi-experimental approaches are used to assess the extent to which sorting is confounding inference for the 2007 cohort of 16 year old school leavers in the National Pupil Database (NPD). The presence of unmeasured pupil characteristics that might be correlated with school type is explored through the use of administrative data that extends beyond the standard NPD controls used in CVA calculations and also through the richness of the Longitudinal Survey of Young People in England. The second section of analysis attempts to use natural variation in the assignment of school autonomy status to assess the effectiveness of (former grant-maintained) foundation schools by comparing the set of schools that just did, and just did not, win a vote of parents to become a grant-maintained schools in the mid-1990s, and thus are usually foundation schools today. In other words, rather than compare progress of pupils at autonomous schools with those at all community schools, this technique evaluates policy by drawing on alternative counterfactual schools.

2. The policy rationale for autonomy

There have been consistent themes running through school autonomy legislation, including the grant-maintained schools policy, City Technology Colleges and the Academies programme. The first claim is that autonomy puts in place both the incentives and the capabilities for substantial improvements in school efficiency. This might be possible because autonomous schools are given control over financial, building and staffing decisions and this proximity of school senior

management teams and governors to the impact of decision-making might lead them to more efficient resource usage, for example in the timing and management of spending on technological projects. Their ability to redeploy saved resources within their own school might incentivise them to use resources more efficiently than community schools.

Secondly, politicians have argued that grant-maintained schools, City Technology Colleges and now Academies would increase parental choice through diversity of school provision and this would be a potential route to greater allocative efficiency in the sense that the schooling system would produce types of schooling that are most desired by parents. However, it is largely agreed that regulations combined with restrictions imposed by the National Curriculum have prevented most autonomous schools from pursuing a particularly distinct curriculum or identity (e.g. Bush et al., 1993; Sherratt 1994).

Finally, it has been suggested that the policies of school autonomy are a catalyst for improvement in standards across the system as these schools provided a competitive threat to their neighbouring schools (or alternatively induced local authority officials to work harder to ensure other schools did not want to leave their control). This direct competition effect, whereby increased competition for (certain) pupils induces neighbouring schools to increase effort directed at exams (thereby improving league table position and making the school more attractive to parents), has been extensively studied in English data with few finding substantive evidence for their impacts (e.g. Allen and Vignoles, 2009; Clark, 2009; Gibbons and Silva, 2008; Gibbons et al., 2008).

Foundation schools

Foundation status for schools arose from legislation in 1998 that brought to a close the grant-maintained schools policy, a relatively radical experiment in school autonomy that gave one-in-six secondary schools independence from Local Education Authority (LEA) control between 1989 and 1997. Grant-maintained schools were owned and managed by their governing bodies, receiving funding directly from the Department for Education which in turn recouped the cost from the

former LEA's revenue support grant. They were completely independent of the former LEA and accountable directly to the Department for Education. The grant-maintained school governing bodies dealt with all staff matters, including suspension and dismissal, though staff enjoyed the same pay and conditions as prior to the transfer of contracts. Schools decided to pursue autonomy for a wide variety of reasons, including greater freedoms, higher levels of funding and avoiding LEA-proposed closures or reorganisations (Bush et al. 1993).

The 1998 legislation that created foundation status protected reduced rights to autonomy from local authority control for grant-maintained schools who retained the same control over operations and admissions as before, gradually lost their preferential financial treatment and were required to have some LEA-appointed governors (Anderson, 2000). Most former grant-maintained schools chose to take foundation status (with some returning to Voluntary-Aided status and a few returning to local authority control). This high degree of association between grant-maintained and foundation status is critical to the estimation strategy for dealing with non-random policy assignment in the second half of this paper.

3. Estimation problem

Studies of school effectiveness attempt to judge pupil outcomes, usually exam scores, holding constant fixed pupil and area characteristics that are outside the control of schools. The measure is intended to represent the extent to which processes that take place in the schools such as behaviour policies, strength of governance and the organisation of teaching increase in pupil achievement. Figure 1 shows a conceptual education production function, whereby outside factors such as school context or pupil characteristics both indirectly and directly impact on the production of pupil achievement. The suggestion is that school autonomy is capable of impacting on the nature of school processes that directly lead to greater pupil achievement.

---------- Figure 1 about here ----------

Contextual Value Added (CVA), the current approach used by government to judge school performance, extracts school residuals (or unexplained variation) from a multilevel regression model that includes basic measures of pupil prior attainment and background and school peer quality. It is an attempt to produce a measure of school achievement that is uncorrelated with the background of the pupil intake so that all schools have an opportunity to be labelled as 'effective'. This is clearly a superior approach to judging effectiveness of schools than 'raw' league table position, but does not produce a direct measure of the strength of the school's processes. Where school processes cannot be explicitly measured, any variation in process that is correlated with contextual control variables will be removed from the estimation of 'effectiveness'. For example, if more affluent schools have systematically higher quality teaching, on average, due to access to a more favourable teacher labour market, calculating school performance conditioning on the context of the school will remove much of the variation in outcomes that was attributable to this higher quality teaching. This does not entirely invalidate school effects that condition on contextual variables, but it does mean that they should only be interpreted as meaningful for comparisons between schools with the same context, rather than a valid method for comparing schools with very different social contexts.

Recent British Government policy has been predicated on the idea that autonomy is causally related to pupil performance, yet it is possible that this inference is confounded by alternative explanations. First, that autonomous schools have pupils with unobserved family circumstances that allow them to make more progress in secondary school. Second, a reverse causation argument that effective schools chose at some point in the past to become autonomous schools and that these differences in effectiveness have simply persisted today.

More formally, we believe there are a set of social characteristics of pupils and schools that affect (i) the chance that a school has become autonomous; (ii) the chance that a child attends an autonomous school; and (iii) pupil GCSE achievement at the school:

$$\text{P(school } j \text{ is autonomous)} = g(r_j, v_j)$$

$$\text{P(child } i \text{ attends an autonomous school)} = h(x_{ij}, w_{ij})$$

$$gcse_{ij} = f(autonomous_j, x_{ij}, w_{ij}, r_j, v_j)$$

The set of observed pupil characteristics, x_{ij}, that might impact on school assignment and achievement include the child's prior attainment, their ethnicity, sex, and so on. The set of unobserved pupil characteristics, w_{ij}, could be the income, social, political or religious characteristics of the household, or more intangible characteristics such as the child's motivation and capacity to learn. The observed school characteristics, r_j, that might impact on the designation of autonomous status and pupil achievement might include the current social composition of the area and the pupils in the school. The unmeasured school characteristics, v_j, could include the historical political control of the area, the characteristics of past headteachers, parent bodies or pupils, and so on.

This article uses two approaches to demonstrate the extent to which these two non-random assignment mechanisms – the assignment of the policy to schools, and of pupils to schools – appear to confound inferences about the causal impact of autonomous status. In the first approach, the number of unmeasured pupil characteristics are minimised as far as possible by supplementing the standard CVA approach to measuring secondary school effectiveness with new data. These additional pupil variables come from two sources. Administrative data from the National Pupil Database is stretched to its analytical limits by utilising early attainment data and small area statistics to act as proxies for unmeasured pupil characteristics. The Longitudinal Survey of Young People in England, which provides rich information on the child's home circumstances, supplements this data for a sub-sample of the population.

The second approach deals with non-random assignment of autonomy status through the exploitation of randomness in the probability that a school became grant-maintained in the early 1990s. A school was only successful in gaining autonomous status of fifty per cent of parents agreed to the move in a ballot, so a regression discontinuity design is employed to compare schools that just did, and just did not, win the vote to become a grant-maintained (and therefore mostly now

foundation) school. Neither of these two approaches deals with both assignment problems simultaneously because re-sorting of pupils has taken place in the 15 years since schools became grant-maintained. That said, where school catchment neighbourhoods tend to have particular fixed characteristics that have persisted over the 15 years, the approach does not confound these area characteristics with effectiveness.

4. Data

Data for over a half a million school-leavers is drawn from the 2007 National Pupil Database (NPD). NPD is an administrative annual census of all pupils in state maintained schools with information on each pupil's sex, age in months, free school meals eligibility (FSM), ethnicity (11 categories), special educational needs (SEN, 3 categories) and mother tongue recorded each year from 2002 onwards. This article uses pupils who are in year 11 (age 16) and are sitting GCSE examinations. These core NPD variables are linked to the child's achievement in Key Stage 2 (KS2) tests in English (reading and writing), maths and science at the end of primary school as a measure of attainment prior to secondary school.

The core NPD variables described above are extended for a subset of pupils who can be observed in all six years of school census data. This allows the matching of multiple measures of pupil background variables that are time-variant, such as FSM and SEN. Also, multiple indicators of the levels of deprivation (income deprivation affecting children index – IDACI) in the localised household area are matched via pupil postcodes. These deprivation indices are based on data from the 2001 Census of Population and other administrative sources (ODPM, 2004) and are all imperfect indicators of the child's social background to the extent that they measure average social characteristics of households in the lower super output area (containing an average of 17 pupils in the cohort). Finally, the child's achievement in a series of Key Stage 1 (KS1) tests is included.

Where school peer group control variables are used in regressions to control for school context these are the percentage of pupils who are FSM eligible, the average KS2 score of the cohort and a

measure of the dispersion of the KS2 scores of the cohort. Table 1 summarises the key variables. The main outcome variable used is total points achieved by pupils over their best 8 GCSE or equivalent examinations and is widely judged as a better discriminator of school performance than the alternative measure of 5+ A*-C grades. The GCSE exam is graded from A* (58 points) to G (16 points), with 6 point increments between grades. The typical pupil achieves around 4 Cs and 4 Ds.

---------- Table 1 about here ----------

The 2,963 schools in the dataset exclude grammar, private, special and other non-standard all-ability secondary schools. Grammar schools are excluded because a very large majority of them became grant-maintained, with no clear counterfactual for the regression discontinuity design that follows. Thus, the entire analysis looks at the impact of autonomy on non-selective schools. The school governance status of the secondary schools is categorised as foundation (non-grammar), voluntary-aided (VA) (non-grammar), voluntary-controlled (VC) (non-grammar), Academy or City Technology College. The default school is a community (LEA controlled) comprehensive school, as shown in Table 2. VA schools are long-standing autonomous schools that (almost all) have a religious foundation whereas VC schools are owned by a religious foundation but are local authority controlled. Academies and City Technology Colleges are grouped together because both groups were small in 2007 (the distinction is becoming less important because many City Technology Colleges are now becoming Academies).

---------- Table 2 about here ----------

The second source of data is the Longitudinal Survey of Young People in England (LSYPE), which charts the progress of a cohort of about 15,000 young people who were initially contacted at age 13/14 in 2004, and are being followed up every year into their mid-twenties. Data analysed here are mostly drawn from the first wave of interviews of the young person and their main parent in 2004. It is supplemented by responses to questions by the young person in wave 3 at age 15/16 and by attainment data and detailed information on type of school attended from the NPD. The timing of

the interviews is less than ideal for these research purposes because we would prefer measures of parental and pupil characteristics prior to secondary school where contamination by the treatment is impossible. For this reason, care is taken to choose variables that reflect reasonably fixed characteristics of the household that are unlikely to be determined by the schooling system.

The LSYPE dataset allows superior analysis of the impact of school type on pupil performance because it has good measures of the social background of families, including indicators of household income, parental education levels and occupational class (see Table 3). It is a survey with moderate levels of non-response on certain variables such as those asked of a second parent. Where non-response is significant the variables are not used in this analysis, except for the household income variables on which a single imputation is performed.

---------- Table 3 about here ----------

The final source of data used in this article is the grant-maintained schools database, which is used to extract voting data from the 1990s for all schools who held a parental ballot to opt-out of local authority control. The database was compiled by the Department for Education in the 1990s and is now archived in the National Digital Archive of Datasets. It provides details of all grant-maintained status ballots taken by schools, and the outcomes of these ballots. It also gives details of major changes to the status of grant-maintained schools, such as requests to change admissions policy or introduce a sixth form. This data is matched to the current administrative data, using school names and postcodes to identify the equivalent current school where identifiers have changed. Table 4 shows the summary characteristics of schools who took a grant-maintained vote, compared to those that did not.

---------- Table 4 about here ----------

5. Dealing with pupil assignment with unmeasured characteristics

This analysis attempts to demonstrate the extent to which pupil characteristics not included in CVA calculations directly impact on both pupil achievement and the probability of attending an

autonomous school, thus confounding CVA-based judgements about the relative performance of different school types. The first part of the analysis extends administrative data for the subset of pupils for whom additional background characteristics are available. The second section supplements administrative data with detailed information on pupils from the Longitudinal Survey of Young People in England to assess the extent to which there are pupil characteristics not measured in the National Pupil Database (NPD) that are correlated with both achievement and type of school attended.

Extending the National Pupil Database

Foundation schools outperform local authority controlled community schools at GCSE in most statistical analysis of NPD. Table 5 shows the key output from four nested regression equations (all clustered to allow for unobserved homogeneity within schools). Capped GCSE score is regressed on four school type variables (with community schools as default) and in the first specification there are no other control variables. The coefficients show that raw GCSE outcomes are about 14 points (or 0.13 s.d.) higher at foundation schools compared to community schools. This means the average child achieves one grade better in two or three of their best eight subjects at these schools, and this difference is statistically significant at the 1% level.

In the second specification a basic set of NPD control variables are added. These include KS2 test scores in separate core subjects, which are responsible for a explaining a large proportion of the variance in the outcome measure, basic pupil indicators of sex, SEN, FSM, ethnicity and age in months. The foundation school coefficient is still statistically significant at the 1% level, but has shrunk to about six points or one grade better in one of eight subjects.

The third specification mirrors the CVA calculation, where based school peer group measures of the FSM and KS2 attainment profile of the school are added. Additional controls indicating pupil mobility (whether the child has moved schools) and small-area deprivation statistics for the child's

home address are included. The advantage of foundation schools is now just two points, or one-third of the grade in one subject, but is still significant at the 5% level.

The magnitude and statistical significance of the foundation school coefficient is little changed by re-running the regression for the sub-group of pupils for whom there are full administrative data from the year 2002 to 2007. This is clearly a non-random group of pupils since they must have been in the English state school system for this entire period, but in this case the selection does not alter inference. At face value, this type of analysis has been used to suggest that sending a child to an autonomous school is beneficial, although the positive impact is small. Extending the NPD to its limits does nothing to change this inference.

---------- Table 5 about here ----------

Longitudinal Survey of Young People in England

A set of household characteristics that we believe are important predictors of child outcomes are collected in the LSYPE and so we can use this dataset to see whether richer pupil background information changes our inferences regarding the impact of school autonomy. Table 6 first replicates basic regressions using the NPD variables in the smaller LSYPE sample to ensure there are not major sampling problems. The estimates of the impact of foundation schools are of a similar magnitude in this sample and are positive and statistically significant. The large number of LSYPE control variables, including parental education, occupation and income, shrink the estimated coefficient on foundation schools to less than 2 points, which is still positive but is now statistically insignificant, so in this sample we reject the hypothesis that foundation status has a causal impact on pupil achievement. LSYPE is a rich dataset yet lacks data on characteristics such as the intrinsic motivation of the child that might predict attainment, but in this circumstance it has proved sufficient to question the validity of CVA-type estimates showing an association between GCSE achievement and school autonomy.

---------- Table 6 about here ----------

6. Dealing with policy assignment on unmeasured school characteristics

The previous section suggests that evidence for the superior performance of foundation schools is weaker than first appears because pupil characteristics that explain both school type attended and attainment are not included in CVA estimates. That said, point estimates on the foundation school coefficient are consistently positive in the samples used. However, even if a positive and statistical significant impact of foundation schools had been consistently estimated, this would not be sufficient to assert causality between the policy of autonomy and achievement because a reverse causation could hold whereby particular types of schools or areas chose to pursue autonomous status.

This section investigates the impact of acquiring foundation status on school performance by exploiting natural policy variation in the acquisition of autonomy. As explained earlier, most foundation schools acquired their autonomous status through the grant-maintained schools legislation of the 1990s and so we can use natural variation in the assignment of this status to analyse the impact of being a foundation school today. The grant-maintained schools policy was very controversial, being opposed by LEAs of all shades; many Anglican and RC churches; some Department for Education officials; large areas of the press; and teacher unionists (Sherratt, 1994). So to legitimise the policy, the government required the Governing Body of schools wishing to acquire grant-maintained status to pass a resolution proposing that an election be held; then win a majority vote of the parents of current pupils. About two-thirds of the c.850 secondary schools who took this vote gained over 50 per cent of the parental vote and thus became grant-maintained schools. Table 7 shows the number of schools for each 2007 governance type that (i) won their vote; (ii) lost their vote; and (iii) never took a grant-maintained schools vote.

---------- Table 7 about here ----------

The fact that a vote of parents was required to become a grant-maintained school provides an important identification strategy for evaluating the policy. There is a clear concern that those

schools taking the vote are a non-random selection of all schools in England, thus any improvement in test scores might be attributable to systematic unobservable characteristics of this group of schools. However, under assumptions discussed in the next section, the schools that just lost the vote to gain grant-maintained status (and so are still community schools today) can be used as the policy counterfactual to those who just won the vote.

A regression discontinuity design (RDD) allows us to identify the effect of becoming a grant-maintained school using the schools that lost their parental vote as the control group. The dichotomous treatment of autonomy ($autonomy_j$) is a deterministic function of the percentage of parents voting yes to grant-maintained status in school j ($vote_j$), with treatment assigned to vote shares greater than 50 per cent. We cannot use the observed differences in outcomes between our groups to infer the impact of treatment on the school's GCSE performance, i.e. $E[gcse^1\text{-}gcse^0]$, because we suspect that a set of (observed or unobserved) covariates such as the affluence or political persuasion of the area, x_j, alters both the school's probability of achieving autonomy status and GCSE outcomes:

$$gcse_j = \beta autonomy_j + g(x_j) + \varepsilon_j$$

$$vote_j = f(x_j) + \upsilon_j$$

The RDD assumes schools near the threshold of 50 per cent are likely to be similar and thus comparable, providing some minimal continuity assumptions for identification are met (see Hahn et al., 2001, for details). It estimates a weighted average treatment effect for the entire population, where the weights are the probability that the school draws a vote share near 50 per cent (Lee, 2005b). This means we can infer little about the potential effects of grant-maintained status for those schools who achieved very low or high vote shares, e.g. 10% or 90%, and indeed for those who did not take the vote at all. This observation that the RDD does not identify the average treatment effect is important because schools taking the grant-maintained vote do not have the same characteristics as other schools, as shown in Table 4. In this sense it is not a perfect substitute for a

randomised experiment across all schools. Even for the group of schools near the 50% vote share there are a set of complicating factors discussed here that determine whether the identification strategy is valid (for a more detailed discussions see Allen, 2008). These complicating factors arise because of (i) the circumstances that led to schools taking a vote; (ii) the capacity of the school to affect its own vote outcome; (iii) the impact of losing the vote; (iv) the ability of a school to re-ballot parents in the event of a lost vote; and (v) a limited number of schools gaining a vote share close to the discontinuity.

(i) The role of the pre-test

The characteristics and GCSE scores prior to treatment of schools taking the grant-maintained vote can be used to partially test the internal validity of the RDD (see Hahn et al., 2001; Lee, 2005a). Data from the Annual Schools Census (presented in Clark, 2009) shows that the mean GCSE scores prior to treatment for vote winners versus losers *just* pass this pre-test of equality at the 5% level on the vote share of 15-85%. Interestingly, the difference between means is consistently negative across all possible chosen groups, i.e. vote winners had lower GCSE performance prior to treatment than vote losers. This is somewhat surprising since intuitively we assume that vote winners were more likely to be in affluent areas without political opposition to grant-maintained status. However, a significant proportion of vote winners were blighted by closure or re-organisation threat (Fitz et al., 1993). One implication of this is the functional form of the estimated RDD – whether the dependent variables is the change in GCSE scores of the new level of GCSE score – can make a substantive difference to short-run estimates.

(ii) Non-random self selection

The grant-maintained schools parental vote represents an unusual application of the RDD because schools have some influence over their vote, with potential non-random self-selection. Specifically, the vote share obtained by the school will be dependent on the headteacher's persuasiveness and campaigning effort (and even parents' perception of the benefit of the treatment), so that, on

average, those who receive the treatment of winning the vote (*vote*≥50) could be systematically more talented or ambitious than those who lost their vote (*vote*<50). Lee (2005a) shows that provided there is some random chance error component to the vote achieved, the treatment can be thought of as statistically randomised around the 50% mark. McCrary (2008) adds a pre-test that should be carried out where agents are able to manipulate the assignment variable, as is the case with headteachers and vote share. The important insight of his test of manipulation is that we should not be able to see significant 'bunching' of observations that just pass the assignment threshold. This test is passed in the grant-maintained schools database, with no unexpectedly large number of schools achieving vote shares between 50 and 55 per cent.

(iii) Independent causal effect of vote share on school performance

In identifying the causal effect of grant-maintained status on school GCSE performance, we must assume that the random draw of vote share does not itself have an impact on the outcome, except through its impact on treatment status (Lee, 2005a). That is, while vote share is allowed to be correlated with GCSE outcomes in the population, the vote is not permitted to have an independent causal impact on outcomes for a given school. There is a plausible argument that in our case the vote share does have an independent causal impact on schools. A school that wins its grant-maintained vote may experience a 'euphoria effect' that temporarily increases staff motivation, resulting in effort directed at improving test scores. Alternatively, a headteacher who wins a controversial vote might experience an increase in respect from staff, allowing them to unite teachers in pursuing exam-orientated goals. Similarly, the school management who loses their vote may well perceive the lack of support for their proposal as a vote of confidence in the school more widely. This would be de-motivating, and may even result cause some vote-losing headteachers to leave their jobs. The consequence of this is that we may overstate the short-run impact of autonomy, and there is some evidence that vote losers did indeed underperform non-vote taking schools in the mid-nineties (Allen, 2008).

(iv) Fuzzy discontinuity

There is a more serious issue in the grant-maintained votes data concerning schools that lost their first vote to become grant-maintained, but that went on to hold subsequent votes, which they won. We use the first ballot win as our assignment to treatment variable since we think it is this vote share that reveals the underlying characteristics of the school, but it is the final ballot that determines the long-run treatment status of the school. Figure 2 shows that 25 of the 233 schools losing their grant-maintained vote went on to hold second or third votes of parents, which they eventually won, thus becoming grant-maintained schools.

---------- Figure 2 about here ----------

There are several possibilities for dealing with this problem in the data. We can use the first vote data in a 'sharp' RDD, as described earlier, and interpret results as 'Intention to Treat' (ITT) estimates (Angrist et al., 1996). However, in this case it is not clear the ITT estimates are the ones we want since we are interested in the effects of school autonomy and not the effects of taking the grant-maintained vote. Alternatively we can use the outcome of a first vote (WIN_j) as an instrument for the outcome of the final vote, known as a 'fuzzy' RDD:

$$gcse_j = \beta_0 + \beta_1 autonomy_j + k(vote_j) + \varepsilon_j$$

$$autonomy_j = \alpha_0 + \alpha_1 WIN_j + h(vote_j) + \upsilon_j$$

This identifies the Local Average Treatment Effect (LATE), which in this case would be the effect of receiving the treatment for schools who won the first vote (and had an expected vote share close to the discontinuity). Both the ITT and IV estimates are reported in the results that follow and coefficients are generally quite similar since few schools held a second vote.

(v) Efficiency versus bias trade-offs

In an ideal application of an RDD we would want to estimate a conditional expectation function at vote=50% since this is not heavily model dependent (Lee, 2005b). This limits approach is contingent on having a great deal of observations very close to the discontinuity, yet just 60 schools had first

vote outcomes between 45 and 55 per cent. This is likely to be an insufficient number of schools to identify an impact, given predicted effect sizes, so risks of Type II statistical errors are high.

The standard parametric approach to dealing with this is to estimate the effect of winning the grant-maintained vote versus losing over a wider band of vote share values (e.g. votes between 15 and 85%) and use linear approximations to generate simple estimates of the discontinuity gap. For example, the dependent variable could be regressed on the vote share, separately on each side of the threshold. This parametric form, which is common in the RDD literature, exploits more data than the use of a narrow band, and can therefore be more efficient. It is also possible that it generates less biased estimates of the true conditional expectation function at the vote=50 threshold than a simple difference in means on a narrower band, where the true function has a non-zero slope. However, the critical assumption is that the parametric regression function used for extrapolation is correctly specified (Lee, 2005b). In our case, we have no *a priori* evidence that vote share should be a linear function of exam score growth. Indeed, this particular regression discontinuity is particularly unusual because it is not entirely clear how vote share should enter the education production function at all: we are quite vague about the unobserved characteristics it proxies.

There are three quite separate sets of characteristics that are unmeasured in the education function and that the vote share may therefore proxy. First, vote share reflects the effort put in by the headteacher and Governing Body to win the vote and it is possible, for example, that heads highly motivated to become grant-maintained are also highly motivated by league table position. Second, vote share reflects the political attitudes of the parental body, and this is correlated with socio-economic background and therefore the academic performance of their children. It could also reflect the degree of confidence that parents have in the school's headteacher more generally, or even their perceived belief in the capacity of the school to benefit from the treatment. Third, vote share reflects the external circumstances the school faced at the time, in particular whether or not it expected to be closed or re-organised in the near future.

That said, there is less concern about bias when estimating a RDD with a large set of covariates to account for differences around the threshold, thus reducing bias on estimates of school effectiveness (and without dimensionality problems since the number of pupils is very large) (Frölich, 2007). These control variables also account for a large proportion of the variance in the outcome GCSE variable, thus increasing the precision of estimates and enabling statistical significance to be potentially achieved for the relatively narrow band of vote winners and losers.

Short-run effects of the grant-maintained schools policy

Clark (2009) applies the RDD to estimate the effect of winning the grant-maintained vote on school GCSE performance for between one and eight years after the change in school status. School GCSE performance is measured as the percentage of pupils gaining 5 or more GCSEs at grades A*-C at age 16 (the only available metric in datasets available for the relevant time period). Clark finds a moderate positive effect of grant-maintained status in the order of a one quarter of a standard deviation (or 4-6 percentage points) change in school performance after two years.

Clark finds this positive effect of grant-maintained status to be persistent over the eight years following the change in school status, but is not able to control for changes in sorting (either through school exclusions or more importantly through school admissions) due to a lack of pupil-level administrative data for this period. This may be important because there was a very sharp increase (as much as three-fold) in the number of permanent exclusion made by schools in the early 1990s, which Gillborn (1996) attributes to increased school competition and the publication of league tables. Clark's positive finding of the impact of grant-maintained status contradicts Levačić and Hardman (1999) who do not employ an RDD, but instead carry out a difference-in-difference analysis of the change in the performance of grant-maintained schools compared to LEA schools from 1991 to 1996. They agree with Clark that on a straightforward comparison of GCSE examination performance, the rate of improvement was higher for grant-maintained schools. However, when they added school control variables they found that this apparently superior performance could be

attributed to having falling proportions of socially disadvantaged students in grant-maintained schools.

Long-run effects of the grant-maintained schools policy

There are difficulties with using data over 15 years after a policy is introduced to identify long-term effects, even where the original policy intervention had a randomised element. First, sorting of pupils across schools, neighbourhoods and even cities will have occurred in this time period. Indeed, theory predicts this is a likely outcome of the intervention. This makes it more difficult to distinguish between school effectiveness and unobserved pupil characteristics. However, since 2002 the National Pupil Database has given us information on where pupils live, enabling us to describe the nature of the sorting effects that control over admissions has on the allocation of pupils across schools. Allen (2007) explores the relationship between residential sorting, school sorting and the presence of foundation schools in area, finding that foundation schools are associated with high levels of post-residential sorting. Figure 3 plots the proportion of pupils eligible for free school meals in the cohort of 2007 school-leavers, grouped by the proportion of parents voting yes to grant-maintained status in the 1990s. It confirms that vote winning schools do now have marginally lower FSM proportions than the vote losing schools that remained as local authority controlled schools, so this should be accounted for when comparing the GCSE achievement of these schools in 2007.

---------- Figure 3 about here ----------

Figure 4 plots the average 2007 capped GCSE point score for schools (grouped to nearest 5%) who took a grant-maintained vote in the 1990s. It shows that there does not appear to be a clear discontinuity in 2007 exam performance between non-grammar schools that did, and did not, win the vote. However, the vote winning schools close to the 50% discontinuity do have a slightly higher capped GCSE score than the vote losing schools.

---------- Figure 4 about here ----------

The visual examination of the recent GCSE data for vote taking schools does not control for sorting of pupils that has taken place since the policy was in place. Table 8 shows the RDD results for a series of regressions of 2007 capped GCSE point scores on whether the grant-maintained vote was won or not. It does this first for schools whose parental yes vote was between 35% and 65%, then for schools whose vote share was between 15% and 85%: the first set of results are less likely to have biased point estimates, but they have fewer observations so less precision. On each occasion both the simple ITT coefficient (whether the school won the first vote) and the two-stage IV coefficient (instrumenting grant-maintained status on the first vote share result) are reported.

The first column of estimates shows that the raw differences in GCSE point scores between vote winners and losers is around 5 points (or one grade difference in one of eight subjects); this is not statistically significant. When full pupil characteristics and peer group variables are added, the gap between winners and losers is actually slightly negative and is again not statistically significant. This suggests that there is no evidence that a policy of school autonomy produces more effective secondary schools. It is perfectly possible that foundation schools are effective schools, but if they are then so are schools that remained in LEA control because they lost the grant-maintained vote. If so, this tells us more about the type of schools that elected to hold a parental vote than the causal effect of a policy, per se. More likely given the analysis in the first part of this paper, there is no genuine difference in the effectiveness of LEA controlled and autonomous schools, with apparent effectiveness of foundation schools attributable to unmeasured characteristics of pupils in these schools today.

---------- Table 8 about here ----------

7. Conclusion

Education policy reforms are usually made without a randomised element so that quantitative researchers struggle to find a good counterfactual for the policy not having taken place. This undermines our ability to draw headline conclusions about whether the policy works. Analysis of

the long-run impact of changes in school governance face two sorting problems: the policy is not usually randomly assigned to areas and pupils start to re-sort across schools in response to the policy. So there is a risk that inference is confounded because schools of a particular level of effectiveness acquire a change in governance, or because particular types of families are attracted to autonomous schools.

This paper has investigated these methodological problems by examining the likely causal impact of foundation school status is on pupil achievement at GCSE. This long-standing policy of autonomy is not particularly radical but gives schools important control over capital, staffing and other operational decisions. It was chosen to analyse because it might inform stable state impacts of autonomy and because the manner of its inception created a control group of schools that failed to acquire the status.

The analysis demonstrates the magnitude of the problem of non-random assignment of pupils to schools using extended pupil background variables from NPD and LSYPE to control for factors that determine both school assignment and pupil achievement. It shows there is little evidence that foundation schools outperform community schools once pupil background characteristics are fully accounted for. The problem of non-random assignment of policy to areas is tackled using the grant-maintained schools vote of parents which created a discontinuity whereby apparently similar schools just did, or did not, receive the treatment. It shows foundation schools that gained their autonomous status as a result of 'just' winning the parental vote perform no better in GCSE examinations than community schools who 'just' lost their parental vote. We should conclude from this that the introduction of grant-maintained status may have led to quite substantial improvements in pupil achievement in the short-term, but these have no persisted to today.

The main purpose of this study has been to showcase methods for overcoming non-random assignment in natural policy settings and we should be cautious in interpreting the substantive findings to autonomy policies more widely. For example, it is perfectly possible that the Academies

programme is successful in raising standards through independence from local authority control since it is far more radical, though early impact evaluations suggest this has not been the case so far (Machin and Wilson, 2008). Similarly, faith schools may well have real characteristics that benefit their pupils beyond secular autonomous schools and there are papers that discuss the extent this is likely to be true using techniques to overcome unobserved pupil characteristics (Allen and Vignoles, 2009; Gibbons and Silva, 2006). Most of the studies mentioned above, taken with this one, suggest that any new policies that give schools autonomy without other major institutional changes are unlikely to lead to sustained improvements in pupil exam performance.

References

Allen, R. (2007). Allocating pupils to their nearest secondary school: The consequences for social and ability stratification, *Urban Studies*, 44:751–770.

Allen, R. (2008) *Choice-based secondary school admissions in England: Social stratification and the distribution of educational outcomes* (PhD Thesis, University of London, UK). Retrieved from http://eprints.ioe.ac.uk/632/.

Allen, R. and Vignoles, A. (2009) *Does school competition matter for pupil achievement? The case of faith schools in England* (DoQSS working paper No. 0904). Retrieved from repec.ioe.ac.uk/REPEc/pdf/qsswp0904.pdf

Anderson, L. (2000) Farewell to grant maintained status: the future of self-governing schools. *School Leadership and Management*, 20(3):371–385.

Angrist, J., Imbens, G. W., and Rubin, D. B. (1996) Identification of causal effects using instrumental variables, *Journal of the American Statistical Society*, 91:448– 458.

Besley, T. and Ghatak, M. (2005) Competition and incentives with motivated agents, *American Economic Review*, 95(3)616-636.

Bush, T., Coleman, M., and Glover, D. (1993) *Managing Autonomous Schools: The Grant-Maintained Experience*, Paul Chapman Publishing Ltd, London.

Clark, D. (2009) The performance and competitive effects of school autonomy, *Journal of Political Economy*, 117(4)745-783.

Department for Education and Skills (2007). GCSE and equivalent results in England, 2005/06. DfES Statistical First Release, Ref: 01/2007: http://www.dfes.gov.uk/rsgateway/DB/SFR/s000702/index.

DiNardo, J. and Lee, D. S. (2004) *Economic impacts of unionization on private sector employers: 1984-2001*, NBER Working Paper, 10598.

Fitz, J., Halpin, D., and Power, S. (1993b) *Grant-maintained schools: Education in the market place*, Kogan Page Limited, London.

Frölich, M. (2007). *Regression discontinuity design with covariates*. IZA Discussion Paper, 3024.

Gibbons, S. and Silva, O. and Machin, S. (2008) Choice, competition and pupil achievement. *Journal of the European Economic Association*, 6 (4). pp. 912-947.

Gibbons, S., Silva, O., and Wilson, J. (2008) Urban density and pupil attainment. *Journal of Urban Economics*, 63(2):631–650.

Gibbons, S. and Silva, O. (2006) Faith schools: better school or better pupils, CEE discussion paper, London School of Economics, CEEDP0072.

Hahn, J., Todd, P., and van der Klaauw, W. (2001) Identification and estimation of treatment effects with a regression-discontinuity design, *Econometrica*, 69:201–209.

Lee, D. S. (2005a) Randomized Experiments from Non-random Selection in U.S. House Elections, UC Berkeley.

Lee, M.J. (2005b) *Micro-econometrics for policy, program and treatment effects.* Oxford University Press, Oxford.

Levačić, R. and Hardman, J. (1999) The performance of grant maintained schools in England: an experiment in autonomy, *Journal of Education Policy*, 14:185–212.

Machin, S. and Wilson, J. (2008) Public and private schooling initiatives in England, in Chakrabarti, R. and Peterson, P.E. (eds) *School choice international: exploring public-private partnerships*, MIT Press, Boston.

McCrary, J. (2008) Manipulation of the running variable in the regression discontinuity design: a density test, *Journal of Econometrics*, 142(2):698-714.

Office of the Deputy Prime Minister (2004) *The English Indices of Deprivation 2004 (revised)*, HMSO.

Sherratt, B. (1994) *Grant-Maintained Status: Considering the Options*. Longman, Harlow.

Tables and figures

Table 1: Key descriptive variables

	Obs	Mean	Std. Dev.	Min	Max
Capped best 8 GCSE scores	577,332	293.27	104.78	0.00	540.00
Five A*-C, including English and maths	577,332	43.89%			
KS2 science score	550,313	4.69	0.69	0.00	6.00
KS2 maths score	550,063	4.38	0.91	0.00	6.00
KS2 English reading score	531,249	4.55	0.82	0.00	5.95
KS2 English writing score	530,878	4.10	0.75	0.00	5.94
Female	577,332	49.44%			
English as an additional language	576,195	10.10%			
White British ethnicity	576,332	81.38%			
Special Education Needs (statemented or school action plus)	560,562	7.88%			
Special Education Needs (school action)	560,562	11.22%			
Free school meals eligibility	560,562	12.85%			
IDACI for home postcode	569,986	0.21	0.18	0.00	0.99

Note: 164 grammar schools are dropped here and for all analysis that follows

Table 2: Governance of secondary schools in dataset

	Percentage of pupils (%)
Community	62.99
Voluntary Controlled	3.09
Voluntary Aided	14.26
Foundation	18.23
CTC/Academy	1.44

Table 3: Summary of additional LSYPE variables

	N	Type
Employment and wealth variables:		
Employment status and type of main parent	14,523	9 categories
Household income (with some imputation for missingness using related variables)	10,810, imputed to 14,569	Continuous
Family receives state benefits such as job seekers allowance or income support	14,308	Binary
Family receives tax credits such as child tax credit	14,308	Binary
Household owns their home	14,386	Binary
Household rents from the council or housing association	14,386	Binary
Household has no (i) computer access; (ii) internet access	14,302; 14,282	Binary
Number of cars in household	14,448	4 categories
Parental education variables:		
Age the main parent left school	14,223	5 categories
Main parent returned to school after leaving	14,294	Binary
Household structure variables:		
Marital status of main parent	14,382	7 categories
Household family structure	14,440	5 categories
Number of children in household	14,409	5 categories
Number of children aged (i) 0-2 years; (ii) 3-11 years; (iii) 12-15 years; (iv) 16-17 years	14,373	2, 4, 3 and 3 categories
Additional child circumstances variables:		
Child was born in the UK	14,351	Binary
Child arrived in the UK in 2000 or more recently	14,351	Binary
Number of schools the child has attended up to age of 13	14,127	4 categories
Number of school moves made at non-standard times	14,127	4 categories

Note: the data is clustered and sample probability weighted

Table 4: Characteristics of vote-taking schools

	Won vote first time		**Lost first vote**		**Never took vote**	
Total KS2 test score	13.62	(2.14)	13.54	(2.20)	13.34	(2.31)
FSM eligibility	8.96%		11.91%		13.90%	
SEN (statement or plus)	6.92%		7.28%		8.18%	
SEN (action)	10.24%		10.62%		11.52%	
White British ethnicity	79.86%		81.68%		81.72%	
English not first language	9.31%		10.17%		10.29%	
Number of pupils	103,043		47,906		426,383	
Number of schools	521		223		2239	
Won subsequent re-ballot			10.37%			

Table 5: School type estimates from NPD regressions

	No controls	NPD controls	CVA	Everything
Foundation	14.12 (1.71) **	6.06 (0.99) **	1.93 (0.93) *	2.25 (1.02) *
Voluntary aided	26.14 (1.94) **	9.29 (1.04) **	6.57 (1.00) **	6.64 (1.10) **
CTC/Academies	-6.52 (8.60) n.s.	4.22 (4.45) n.s.	28.48 (3.91) **	28.77 (4.26) **
Voluntary controlled	16.93 (3.16) **	6.37 (1.91) **	7.41 (2.27) **	6.37 (2.30) **
4 KS2 scores (maths, science, English, total squared)	No	Yes	Yes	Yes
Sex, EAL, SEN, FSM, ethnicity (11), age	No	Yes	Yes	Yes
School peer (FSM, KS2, s.d. of KS2)	No	No	Yes	Yes
Extra background (IDACI and mobility)	No	No	Yes	Yes
Extended NPD (KS2 sub-scores; 4 KS1 scores; multiple SEN, FSM and IDACI)	No	No	No	Yes
R-squared	0.88%	51.30%	53.56%	46.06%
Number of pupils	582,347	547,716	545,568	358,393
Number of schools	2964	2936	2936	2933

Notes: (a) regression estimates clustered for unobserved school homogeneity
(b) **=significant at 1%; *=significant at 5% level

Table 6: School type estimates from LSYPE regressions

	No controls	NPD controls	LSYPE controls
Foundation	20.47 (5.56) **	8.65 (3.48) *	1.88 (2.67) n.s.
Voluntary aided	33.78 (4.75) **	12.13 (3.09) ***	9.59 (2.69) **
CTC/Academies	82.33 (5.19) **	30.53 (14.48) *	31.84 (11.22) **
Voluntary controlled	35.36 (3.15) **	13.86 (9.33) n.s.	4.45 (8.87) n.s.
4 KS2 scores (maths, science, reading, writing)	No	Yes	Yes
Sex, EAL, SEN, FSM, ethnicity (11), age	No	Yes	Yes
School peer (FSM, FSMsq, KS2, KS2sq)	No	No	Yes
Household income and benefits, parental education and employment, family structure, computer/internet/car at home, country of origin, mobility information, home ownership.	No	No	Yes
R-squared	1.81%	50.14%	56.11%
Number of pupils	14,097	13,149	12,260

Notes: (a) weights applied to account for sampling structure
(b) **=significant at 1%; *=significant at 5% level

Table 7: Governance in 2007 of grant-maintained vote winners and losers

	Final grant-maintained ballot won	Final grant-maintained ballot lost	Grant-maintained ballot never taken
Grammar (all governance types)	94	8	70
CTC/Academy	0	0	50
Community (non-grammar)	10	151	1,678
Foundation (non-grammar)	404	13	90
Voluntary-aided (non-grammar)	130	26	334
Voluntary-controlled (non-grammar)	0	10	67
	638	208	2289

Table 8: RDD of vote winners and losers

		No controls	NPD controls	CVA	Everything
V = [35,65]	Vote winners (ITT)	4.92 (4.73)	-0.12 (2.54)	-1.14 (2.30)	-1.95 (2.56)
	Vote winners (IV)	5.83 (5.59)	-0.14 (3.01)	-1.36 (2.72)	-2.33 (3.05)
	R-squared	0.08%	50.04%	51.08%	45.63%
	Number of pupils	46301	42663	42663	27784
	Number of schools	219	218	218	218
V = [15,85]	Vote winners (ITT)	4.65 (3.17)	1.64 (1.84)	0.46 (1.69)	0.14 (1.91)
	Vote winners (IV)	5.23 (3.55)	1.85 (2.07)	0.51 (1.90)	0.16 (2.14)
	R-squared	0.08%	48.49%	49.46%	44.92%
	Number of pupils	118684	108783	108783	71857
	Number of schools	567	566	566	566
Controls	KS2 marks in core subjects	No	Yes	Yes	Yes
	Standard NPD controls	No	Yes	Yes	Yes
	School peer controls	No	No	Yes	Yes
	Extended NPD, incl. KS1	No	No	No	Yes

Note: regression estimates clustered for unobserved school homogeneity
See earlier NPD regressions for full details of control variables included

Figure 1: An education production function

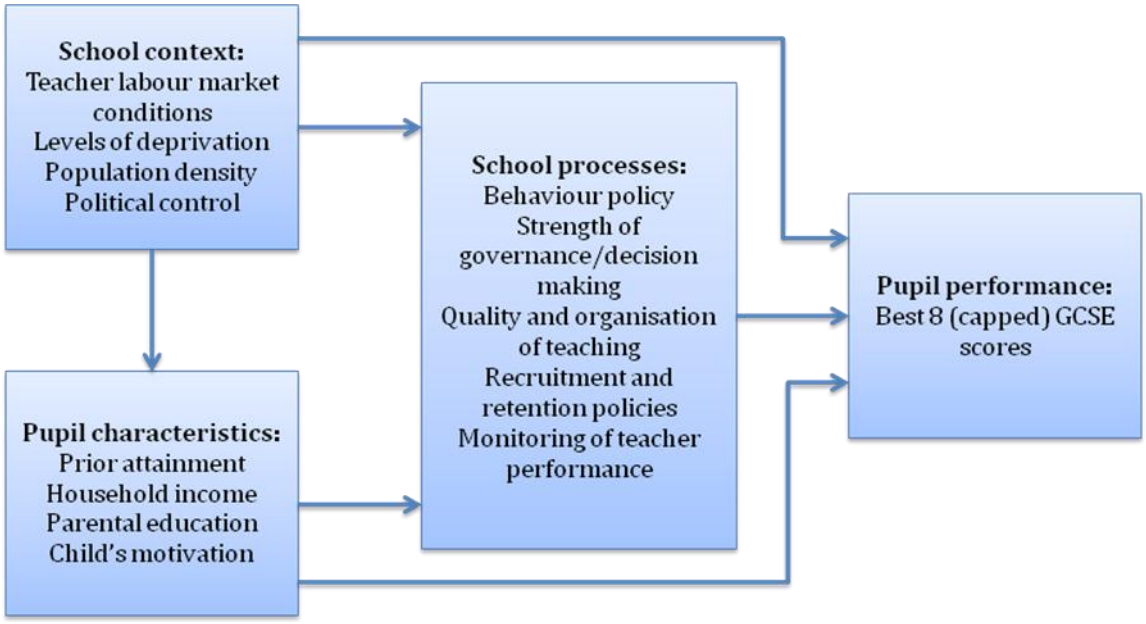


Figure 2: Re-balloting creates a fuzzy discontinuity

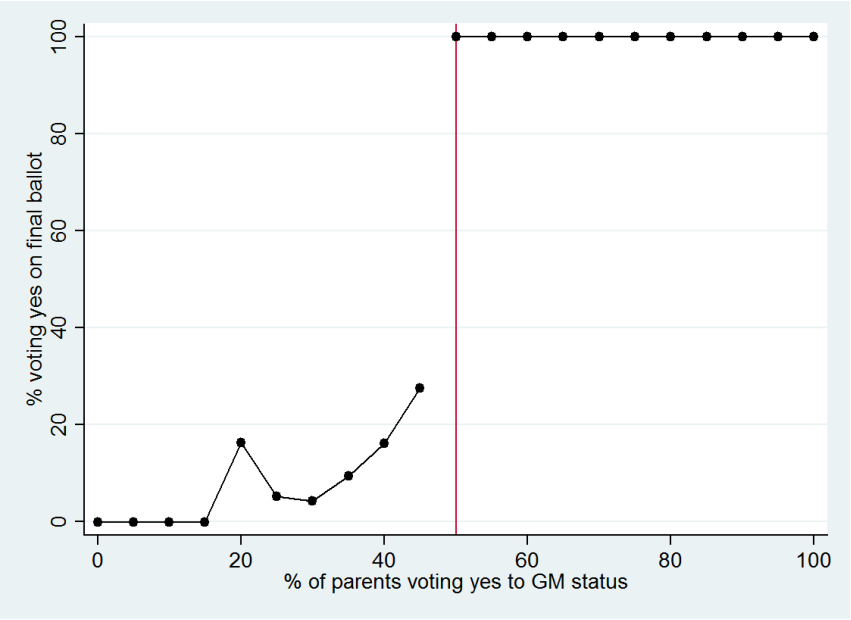


Figure 3: 2007 FSM proportions at vote taking schools

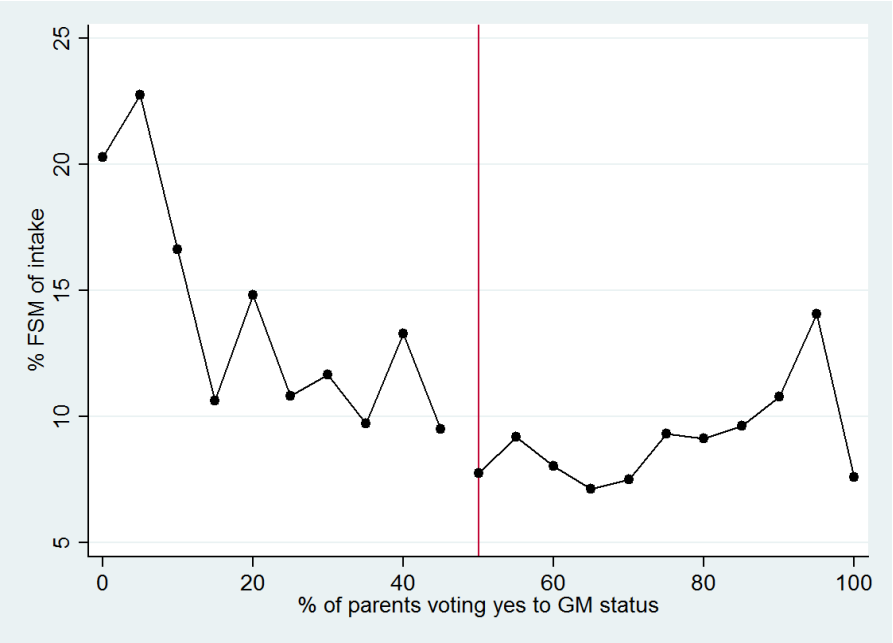
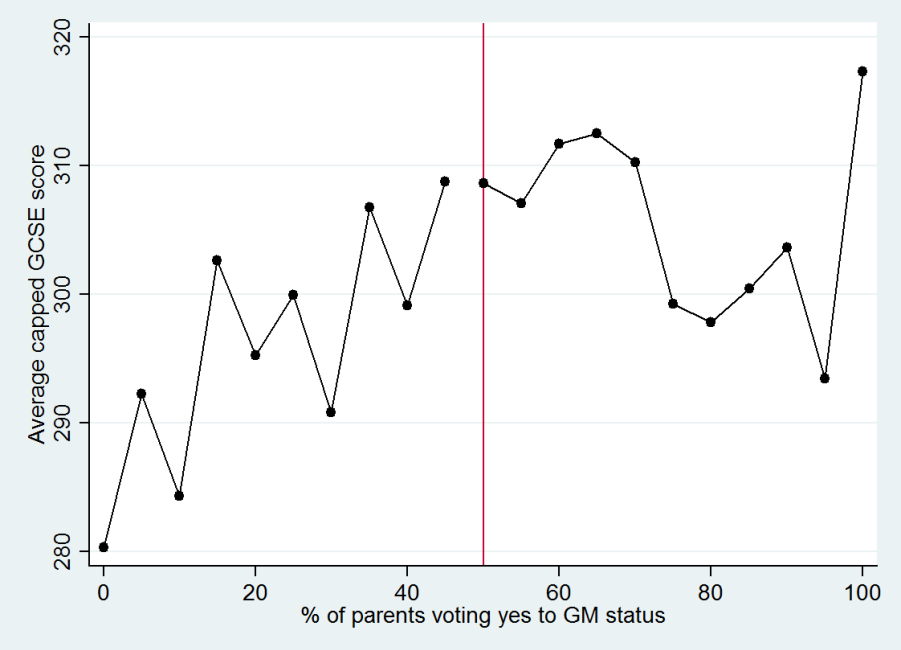


Figure 4: Average capped GCSE score at vote-taking schools in 2007



Note: Charts for alternative outcome measures (e.g. 5+ A*-C) look identical