Institute for Fiscal Studies

Jack Britton
Damon Clark
Ines Lee

Working paper

23/24

Exploiting discontinuities in secondary school attendance to evaluate value added

Economic and Social Research Council

Exploiting discontinuities in secondary school attendance to evaluate value added

Jack Britton[1], Damon Clark[2], and Ines Lee[3]

[1]University of York and Institute for Fiscal Studies
[2]University of California, Irvine
[3]University of York

September 19, 2023

Abstract

We use a new empirical strategy to test various measures of school effectiveness in England. Our approach exploits discontinuities in attendance probabilities that occur at unpredictable distance cutoffs used as tiebreakers in admissions processes for oversubscribed schools. We show that raw, unadjusted test score outcomes of schools are biased estimators of school effectiveness due to sorting of certain types of students into certain types of schools. Controlling for basic background characteristics (but not prior attainment) does not change this result. On the other hand, we cannot reject the hypothesis that simple value added models which adjust school test scores for difference in the prior attainment of their intake produce unbiased effectiveness estimates. This includes "Progress 8", the measure of value added that has been used in England since 2016, suggesting the measure accurately captures the true effect of schools on pupils' GCSE attainment. Adding additional background controls does not invalidate the estimates, but it does not improve precision either. Finally, we combine our unbiased estimator of effectiveness with data on secondary school applications to show that parents often do not put the most effective school in their local area down as their first choice school. This is particularly true for parents from poorer households, suggesting SES gaps in access to good schools could be narrowed through changes to school application patterns.[1]

Keywords: value added models, school effectiveness
JEL Codes: I20, J24, C52

[1]This work has been funded by the Nuffield Foundation. The Nuffield Foundation is an independent charitable trust with a mission to advance social well-being. It funds research that informs social policy, primarily in Education, Welfare, and Justice. It also funds student programmes that provide opportunities for young people to develop skills in quantitative and scientific methods. The Nuffield Foundation is the founder and co-funder of the Nuffield Council on Bioethics and the Ada Lovelace Institute. The Foundation has funded this project, but the views expressed are those of the authors and not necessarily the Foundation. We thank our Advisory Group for their contribution to the work and Laura van der Erve and Louis Hodge for additional research input. This work contains statistical data from ONS which is Crown Copyright. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates. The work was carried out in the Secure Research Service, part of the Office for National Statistics.

1 Introduction

Measuring school effectiveness is important for informing parental choice, for incentivising high teaching standards, for identifying good (and bad) school practice, and for thinking about equality of opportunity in education. It is generally accepted that using raw, unadjusted school outcomes, such as average test scores or the proportion of students passing a certain threshold,[2] are not good measures of effectiveness due to there being large differences in the types of students attending different schools.

Value added models (VAMs) attempt to address this by adjusting school outputs (usually student test scores) for the prior achievement and often additional demographic characteristics of the student body. In England, VAMs of various guises have been used to measure secondary school effectiveness since the early 2000s (see Leckie and Goldstein, 2017, for a review). The most recent of these was introduced in 2016 and is called "Progress 8". This is a very simple VAM that *only* adjusts for prior test scores. The decision to only adjust for prior attainment and not any additional student characteristics has been criticised on the grounds that it is unlikely to satisfactorily control for the way parents select into schools, which may generate inaccurate estimates of effectiveness (Leckie and Goldstein, 2017).

However in practice, we do not know whether Progress 8 is a reliable measure. Clarity on this issue is an important policy issue in the United Kingdom, as Progress 8 is highly visible and influential.[3] It is prominent on school comparison sites and is used to create publicly available school rankings that influence parental choice and that teachers care about. Poor Progress 8 results can trigger inspections from the standards regulator and potentially influence school closure decisions. And according to current teachers, it is increasingly influencing school practices.[4]

Although there are previous studies that have estimated bias in VAMs (Deming, 2014; Angrist et al., 2017; Abdulkadiroğlu et al., 2020; Angrist et al., 2021), the applicability of those findings to the English setting is unclear due to differences in institutional settings, differences in methodological choices, and differences in measures of school inputs and outputs. And in any case, the

[2]For example, in England the share of students who attained at least a C grade in five or more GCSE exams was a commonly reported measure of school performance.

[3]The question is also relevant more generally. One clear advantage of Progress 8 (and similar variants) is that it is not so demanding in terms of data, making it more realistic to introduce similar measures in settings other than in England.

[4]For example, see the Twitter account of Katherine Birbalsingh, the Headteacher of Micheala school in London, which was the highest ranked school in the country for Progress 8, in October 2022, when that year's Progress 8 results were released (she was also Chair of the UK Social Mobility Commission at the time). She highlights her schools success but also encourages other schools to emulate the practices of the high performing schools in the Progress 8 rankings.

previous evidence is highly ambiguous on this issue - while some papers find VAMs to be unbiased, others find the opposite.

The studies listed above all evaluate the validity of VAMs by exploiting admissions lotteries that are used to determine entry to oversubscribed schools in the United States. The basic intuition is that if the VAM is valid, test score gains from winning an admissions lottery should align with the gain predicted by VAM. However, like in most settings, admissions lotteries do not exist in England. In this paper we present a new empirical strategy for identifying causal school effects on test scores in England, which uses discontinuities in attendance probabilities around the edges of unpredictable distance cutoffs used as tiebreakers in admissions processes for oversubscribed English secondary schools. We use this to test whether a range of different measures produce unbiased estimates of school effectiveness.

In English state secondary schools there are often sharp discontinuities in attendance that occur at certain distances from the school to prospective students' homes. Many of these discontinuities are generated by natural barriers, such as rivers or roads, or specified catchment areas in which pupils are close to guaranteed admission. Such discontinuities are unlikely to generate credible experiments, as parents can select to live inside natural barriers or catchment areas in order to attend their preferred school. However, many schools in the country remain oversubscribed once their other selection criteria are applied. And in such cases, distance is commonly used as a final tiebreaker. Since this distance cutoff depends on demand in a given cohort, it changes over time and is not easily predicted, meaning it is not possible to manipulate it to be just inside or just outside the cutoff.

Using a combination of web searches and Freedom of Information requests, we collected data from secondary school admissions brochures published by Local Authorities between 2007 and 2012 on which schools used distance as a tiebreaker and the precise distance used. We then matched this information to individual-level administrative data on all secondary pupils in England. We argue that schools which have sharp attendance discontinuities at the precise distance that corresponds to the brochure cutoff produce plausibly causal identification of the effect of attending that school versus the school that individuals would have attended had they lived on the other side of the cutoff.

We exploit this quasi-experimental variation to test a set of five measures of school effectiveness measures for bias with varying sets of controls methods of construction. We find strong evidence that measures of school performance which do not adjust for prior attainment are biased. However, we are unable to reject the hypothesis that simple VAMs that adjust for prior attainment only produce unbiased effectiveness estimates. This includes the government's headline Progress 8 measure, suggesting it is reliable estimator of the true effect of secondary schools on GCSE results. However we do find some suggestive evidence that controlling more flexibly for prior attainment than Progress 8 does would produce a more reliable measure. Once we do this, adding additional student demographic controls does not improve precision, weakening the case for incorporating additional student demographic controls into the headline measure.

This is a growing branch of the literature that has developed out of research evaluating the validity of teacher VAMs (Kane and Staiger, 2008; Kane et al., 2013; Chetty et al., 2014; Bau and Das, 2020). These papers show that estimates of teacher effectiveness based on quasi-randomisation of students to teachers broadly align with OLS-based estimates of teacher effectiveness, suggesting such measures are unbiased. As mentioned above, we connect most directly to the small number of papers that apply similar tests to evaluate school value added using school admissions lotteries (see Angrist et al., 2022, for a comprehensive review of the evidence). Deming (2014) uses data on lotteries from 118 schools in North Carolina, finding that even relatively simple value added estimates are unbiased using a similar approach to those used in the teacher value added literature. Angrist et al. (2017) formulates a new, more stringent test of value added bias via an overidentification test of the orthogonality restrictions generated by a set of lottery instruments from 29 schools in the Boston area. They find that while the forecast coefficient test suggests that value added is unbiased, their overidentification test rejects the null that VAM estimates correctly predict the effect of randomized admission at ever school with a lottery. They present evidence suggesting that non-charter schools drive this result. Abdulkadiroğlu et al. (2020) implement the tests suggested by Angrist et al. (2017) using 124 lotteries in New York City, concluding that the VAM they use is unbiased for maths scores, but not for the more general PSAT scores. Finally, Angrist et al. (2021) finds unbiased estimates for middle schools in Denver (based on 67 lotteries) and New York City (448 lotteries), but biased estimates for New York City high schools (382 lotteries). This evidence is therefore very mixed, though generally leaning towards VAMs being unbiased.

Our paper contributes to this literature by adding another test of VAM bias in a new institutional setting, exploiting a different source of identification. To our knowledge, the only other papers that test for school VAM bias using identification sources other than lotteries are Ainsworth et al. (2022), which exploits discontinuities in attendance driven by entry requirements based on prior attainment in Romania, and Andrabi et al. (2022), which exploits school switches driven by school closures in Pakistan. Both papers conclude that the value added measures they test are unbiased. As described above, our work is also highly relevant to the policy debate in the UK, where findings from tests in different institutional settings are unlikely to apply.

An additional contribution of our work is to show that our data on the precise distance cutoffs used by schools is crucial for identification. We find that we are unable to credibly increase our sample by searching using for attendance discontinuities using a search algorithm, such as the one used by Hoekstra (2009), even if we restrict our search to being within close proximity of reported brochure cutoffs. This is an important result for those intending to test VAM validity in alternative settings and for researchers trying to exploit attendance discontinuities around the edge of secondary schools to answer other research questions.

In the final section of the paper, we draw upon the finding that we have credible estimates of school effectiveness to investigate the extent to which parents leave value added "on the table" when making their secondary school choices (following a similar approach to Ainsworth et al., 2022). Making use of secondary applications data from 2014, we show that there are large differences between the first preference school of parents and the most effective school in their local area, even when that local area is restricted so that parents could not select a school that is further away than the one their child actually attended. The gaps between the effectiveness of the first choice school and effectiveness of the most effective local school are larger for parents from poorer households than they are for parents from richer households, suggesting SES gaps in school quality could be narrowed considerably if parents were to select on school effectiveness.

The rest of the paper is set out as follows. We discuss VAMs and results from the previous literature in more detail in Section 2. We provide details on the English school system and the data in Section 3 before outlining our empirical framework in Section 4. We show our results and some robustness checks in Section 5. Section 6 presents our investigation of whether value added is left on the table, before Section 7 concludes.

2 Estimating school VAMs and testing for bias

2.1 School VAMs

Value added models (VAMs) commonly take the following basic form:

$$Y_{ij} = \sum_{j=1}^{J} \lambda_j D_{ij} + X_i' \delta + \epsilon_{ij} \tag{1}$$

where Y_{ij} is a school outcomes measures (usually test scores), D_{ij} is a dummy for student i attending school j with school 0 omitted, $\hat{\lambda}_j$ is estimated impact of school j on test scores, or the estimated VA of school j relative to school 0. Finally, X_i is vector of control variables, which must include prior test scores, to be considered a "value-added" model, but often includes additional background variables. The key assumption is that this model is well specified and $\hat{\lambda}_j$ identifies the causal impact of the school.

There are several ways to estimate VAMs, including different X variables (e.g. demographic variables such as free school meal eligibility and ethnicity), different outcome variables (e.g. test scores, university entry or earnings), different specifications (e.g. interacted models) and different ways of dealing with noise (e.g. pooling, shrinkage or random effects).

Progress 8, the headline measure of effectiveness for English secondary schools is not constructed through a regression approach. The details of how it is constructed are given later in the paper, but for the purposes of the following discussion, it is helpful to have the simple VAM from equation (1) in mind.

2.2 Testing for bias using lotteries

Previous studies have tested school VAMs by exploiting admissions lotteries. The basic intuition behind these tests is that if a school VAM is unbiased, the difference in test scores between lottery winners and lottery losers should align with the difference in value added between lottery winners and losers. That is, one would expect $\pi^Y = \pi^V$ from the following:

$$Y_{ij} = \alpha^Y + \pi^Y Z_{ij} + X_i' \Pi^Y + \eta_{ij}^Y \tag{2}$$

$$V_{ij} = \alpha^V + \pi^V Z_{ij} + X_i' \Pi^V + \eta_{ij}^V \tag{3}$$

where Y_{ij} is test scores of individual i attending school j, V is value added (using $V = \hat{\lambda}$ from above), Z is a dummy for winning the lottery, X is a set of control variables, $\alpha^{Y,V}$ are constants and $\eta^{Y,V}$ are random noise.

This can be recast to a two-stage least squares set up:

$$Y_{ij} = a^{Y} + \beta \hat{V}_{ij} + X_i' \Theta^{Y} + \omega_{ij}^{Y} \tag{4}$$

$$V_{ij} = a^{V} + \pi Z_{ij} + X_i' \Theta^{V} + \omega_{ij}^{V} \tag{5}$$

where $\hat{V}$ in (4) is estimated from the first stage equation (5). Here we expect $\hat{\beta} = 1$ since by definition $\hat{\beta} = \frac{\pi^Y}{\pi^V}$. In the literature, $\hat{\beta}$ is referred to as the forecast coefficient.

This idea is straightforwardly extended to a setting where there are K lotteries, each associated with a different school (we include a dummy for each school here and now drop the constant). Let L be a dummy for entering the lottery, we then estimate:

$$Y_{ij} = \sum_{j=1}^{K} \theta_j^{MY} L_{ji} + \beta^{M} \hat{V}_{ij} + X_i' \Gamma^{MY} + \zeta_{ij}^{MY} \tag{6}$$

$$V_{ij} = \sum_{j=1}^{K} \theta_j^{MV} L_{ij} + \sum_{j=1}^{K} \pi_j^{M} L_{ij} Z_{ij} + X_i' \Gamma^{MV} + \zeta_{ij}^{MV} \tag{7}$$

The test is then whether or not $\hat{\beta}^M = 1$. This approach mirrors Angrist et al. (2017), Abdulkadiroğlu et al. (2020) and Angrist et al. (2021). Later, we refer to this approach as the 'multi-IV' approach (hence the "M" superscript) as it treats each of the lotteries as separate instruments for value added. We use this to distinguish from the 'single-IV' approach, which interacts the instrument, Z with $\Delta V = V_{ji} - V_{-ji}$, so we have:

$$Y_{ij} = \sum_{j=1}^{K} \theta_j^{SY} L_{ji} + \beta^{S} \hat{V}_{ij} + \tilde{X}_i' \Gamma^{SY} + \zeta_{ij}^{SY} \tag{8}$$

$$V_{ij} = \sum_{j=1}^{K} \theta_j^{SV} L_{ij} + \pi^{S} \sum_{j=1}^{K} L_{ij} Z_{ij} \Delta V_{ij} + \tilde{X}_i' \Gamma^{MV} + \zeta_{ij}^{SV} \tag{9}$$

The key difference here is that this involves the estimation of one parameter on the instruments, π^S, rather than one parameter for each instrument (π_j^M in the multi-IV approach, equation 7). This approach is similar to the one used by Deming (2014),[5] and the test is now whether $\hat{\beta}^S = 1$.

[5]This approach also requires the inclusion of an additional control for V_{-ji}, which add to X to create $\tilde{X}$.

Table 1 summarises the evidence from four previous papers that have used lotteries to test value added bias. Deming (2014) finds that when average test scores in maths and reading are adjusted for lagged test scores, he is unable to reject the null that resulting estimate of school effectiveness is unbiased (the forecast coefficient is 0.966, while the p value is 0.92) for a subset of middle schools in North Carolina.

Table 1: Literature Summary

	Angrist et al. (2017)		Abdulkadiroglu et al. (2020)	
Setting	Boston	Boston	New York City	New York City
Outcome:	Math	Math	Math	PSAT
School level	Middle	Middle	High	High
Lagged scores	Yes	Yes	Yes	Yes
Background chars.	Yes	Yes	Yes	Yes
Sample selection:		*Exc. Charters*		
Forecast coefficient	0.864	0.549	0.965	0.879
s.e.	[0.075]	[0.164]	[0.038]	[0.048]
p value	0.071	0.006	0.354	0.012
First stage F-Stat	29.6	11.2	6.1	2.3
OverID test p value	0.003	0.043	0.996	0.080
Lotteries	29	24	124	124
N	8718	6162	32131	32131
	Deming (2014)	**Angrist et al. (2021)**		
Setting	North Carolina	Denver	New York City	New York City
Outcome:	Math/reading	Math	Math	SAT Math
School level	Middle	Middle	Middle	High
Lagged scores	Yes	Yes	Yes	Yes
Background chars.	No	Yes	Yes	Yes
Forecast coefficient	0.966	1.120	0.933	0.783
s.e.	[0.342]	[0.106]	[0.041]	[0.064]
p value	0.920	0.275	0.105	0.001
First stage F-Stat	*not reported*	104.0	649.0	240.0
OverID test p value	N/a	0.070	0.186	0.043
Lotteries	118	67	448	382
N	2599	7660	44498	30066

Abdulkadiroğlu et al. (2020) get the same result for test scores in math, but not the PSAT, for which they also reject the null, based on a subset of New York City high schools. Angrist et al. (2017) estimates a forecast coefficient of 0.86 (p=0.07) which is interpreted as a borderline pass of the null. However, crucially Angrist et al. (2017) introduce a more stringent test of the overidentification restrictions in their multiple instrument setup. They find that this test is rejected, meaning that while the forecast coefficient is close to one overall, it is not close to one for some of the school experiments. The authors are able to explain this by showing that once they exclude Charter schools from their analysis, the forecast coefficient *is* significantly different to one. Abdulkadiroğlu et al. (2020) also perform this test, finding a pass for math scores and a borderline pass for PSAT scores. Finally, Angrist et al. (2021) do not reject the null that basic value added models are unbiased for middle schools in Denver and New York, although they present only a borderline pass of the overidentification test in the former case. For New York City high schools, however, they reject the null that the basic value added model is unbiased, as they estimate the forecast coefficient to be 0.78 (p=0.001). They also reject the null for the overidentification test in this setting.

The previous evidence on "conventional" value added models that control for lagged test scores and often additional background characteristics of students is therefore very mixed, highlighting the need for testing VAM bias in individual institutional settings.[6]

3 Institutional background and data

We study English secondary schools, which pupils attend from age 11 to age 16 (also referred to as Year 7 to Year 11). At the end of Year 11, secondary school pupils take the General Certificate of Secondary Education (GCSE) examinations. Prior to entry to secondary school, pupils also take national examinations called Standard Assessment Tests (which we refer to as "Year 6 SATs") in the final of year of primary school.[7] The combination of these tests are used to construct 'Progress 8' which is now the most prominent measure of value added of English secondary schools. These scores are made publicly available, and are therefore a highly visible measure of school effectiveness. They can influence hiring and firing of teachers, school practices, and where schools perform poorly, can trigger inspections.[8]

3.1 Details of student assignment process

Like many countries, English secondary schools do not use lotteries to determine entry. We will therefore exploit alternative features of the school assignment process detailed below.

Parents (or caregivers) apply to secondary schools for their children in January of Year 6, the

[6]Conversely, the literature is unequivocal that using test scores without adjusting for background characteristics does produce biased estimates of school effectiveness. Deming (2014), Abdulkadiroğlu et al. (2020) and Angrist et al. (2021) all get forecast coefficients that are significantly different to one when testing this.

[7]We also have data on "Year 2 SATs" although these exams are not included as inputs to Progress 8.

[8]Schools are also routinely subjected to inspections from the Office for Standards in Education (Ofsted). Poor Progress 8 performance can trigger an inspection - otherwise, schools are inspected roughly every four-five years, depending on the results of previous inspections.

final year of primary school. Pupils are then assigned to schools via a centralised deferred acceptance, Gale-Shapley algorithm. The algorithms works as follows:

1. Assign all pupils to their c^{th} choice school (starting with $c = 1$, the first choice school).
2. For schools where the number of assigned students (N_s) exceeds the school's capacity (S_s), rank pupils by school preferences r_s and keep students for whom $r_s \leq S_s$.
3. Reassign all students for whom $S_s < r_s \leq N_s$ to their $(c+1)^{th}$ choice school.
4. Repeat steps (2) and (3) until all pupils are assigned.

The algorithm crucially depends on r_s, the school ranking of students. These will typically be based on priorities such as:

1. Pupils with special educational needs
2. Pupils with a sibling in the school
3. In some cases, pupils who attended feeder primary school
4. In some cases, pupils with a specific religious affiliation
5. Everyone else[9]

Since these are discrete variables, a tiebreaker is required to sort students in each category. Typically, the distance from the pupils' home address to the school at the time of application is used for this.[10] We refer to a school for which at any point in the algorithm $N_s > S_s$ as being oversubscribed in a given application cycle. For such schools that then use distance as a final tiebreaker to determine entry, there may be discontinuities in the probability of attendance at the school around the specific distance cutoff used. Since the specific cutoff is based on the number of priority students, the number of first preferences for the school, and the rejection rates of local schools, the precise location of these distance cutoffs are impossible difficult to predict ex-ante. This therefore generates pseudo-randomisation of households into or out of the school depending on whether they are just inside or just outside the distance threshold. It is this pseudo-randomisation that we are interested in exploiting in order to test VAM validity.

[9] Importantly, schools are not allowed to select on the students' preference ranking, and in fact do not observe this. In theory this makes the process 'strategy-proof' meaning parents should order their school preferences truthfully.

[10] For some schools, other tiebreakers are used. For example, schools are able to differentiate students based on their evidence of religious affiliation. A very small number of London schools also used test scores as a tiebreaker.

We do not have data on all of the assignment conditions. For example, the Department for Education do not release information on siblings, and we do not know religion. However, we are able to make use of information on specific distance tiebreakers.

3.2 Brochures

We require information on whether schools used a distance tiebreaker *and* the precise distance used. This is necessary because discontinuities in attendance can occur at certain distances from a school for reasons that are predictable, and they are therefore are likely to suffer from selection of parents who choose to live inside or outside the cutoff in order to influence which school their children can attend. For example, a river or a motorway might result in a big drop in attendance probabilities around a desirable school, and parents might select to living on the side of the river or motorway that makes access to the desirable school easier. We therefore do not think it is reasonable to search for attendance discontinuities around all schools without being guided by information on whether brochure cutoffs are used.

Table 2: Summary of brochure data

	No. schools with any cutoff info	Median distance (km)
2007	298	2.94
2008	488	3.29
2009	418	3.33
2010	444	3.01
2011	528	3.37
2012	502	3.51
No. school-year obs	2678	
No. schools	820	

Source: Local Authority Brochures collected by the authors. Year is based on the secondary school cohort the distance cutoff applied to.

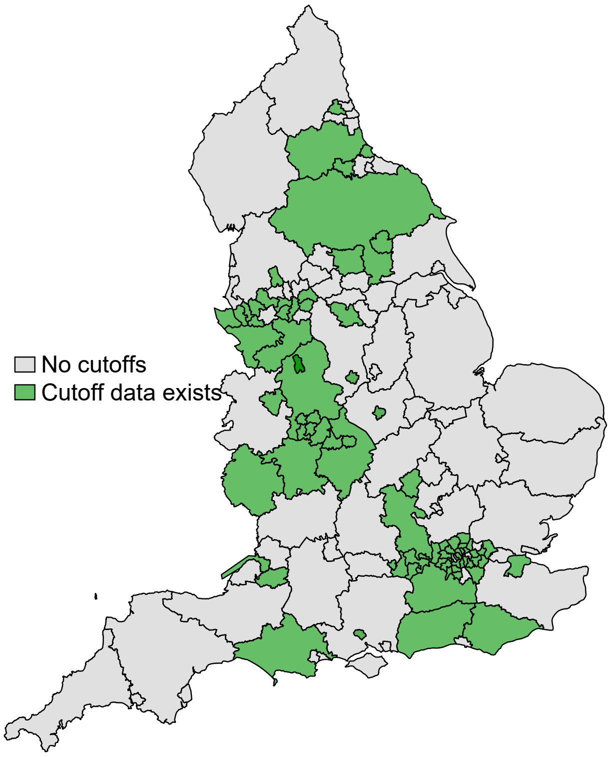
Fortunately for our research design, this information is available, as Local Authorities in England published admissions brochures that included information on the selection processes of secondary schools in their area, whether distance was used by the school as a tiebreaker, and if so, the precise distance that was used.[11] We collected the brochures through a combination of internet searches and freedom of information requests sent to Local Authorities. The basic data that we collected is summarised in Table 2. Across six years of data, we have 2678 precise distance cutoffs from 820 different schools (with many schools reporting cutoffs for multiple years). The

[11]See Appendix A for examples.

years correspond to the year of entry to secondary school, meaning the 298 cutoffs we observe in 2007 apply to the cohort of students which entered Year 7 in September 2007. We observe more distance cutoffs (around 500) in later years as we were able to collect more brochures in the later period. There were approximately 4,000 secondary schools in England during this period, meaning around 20% report using distance cutoffs. The table also reports the median reported cutoff distance in each year, which is around 3km in each year.

These brochures are not a representative sample from across the whole country - Figure 1 shows that the brochures we collected are predominantly from London, the South and the West Midlands. This is not surprising as we expect oversubscription to be used in areas where there is high population density.

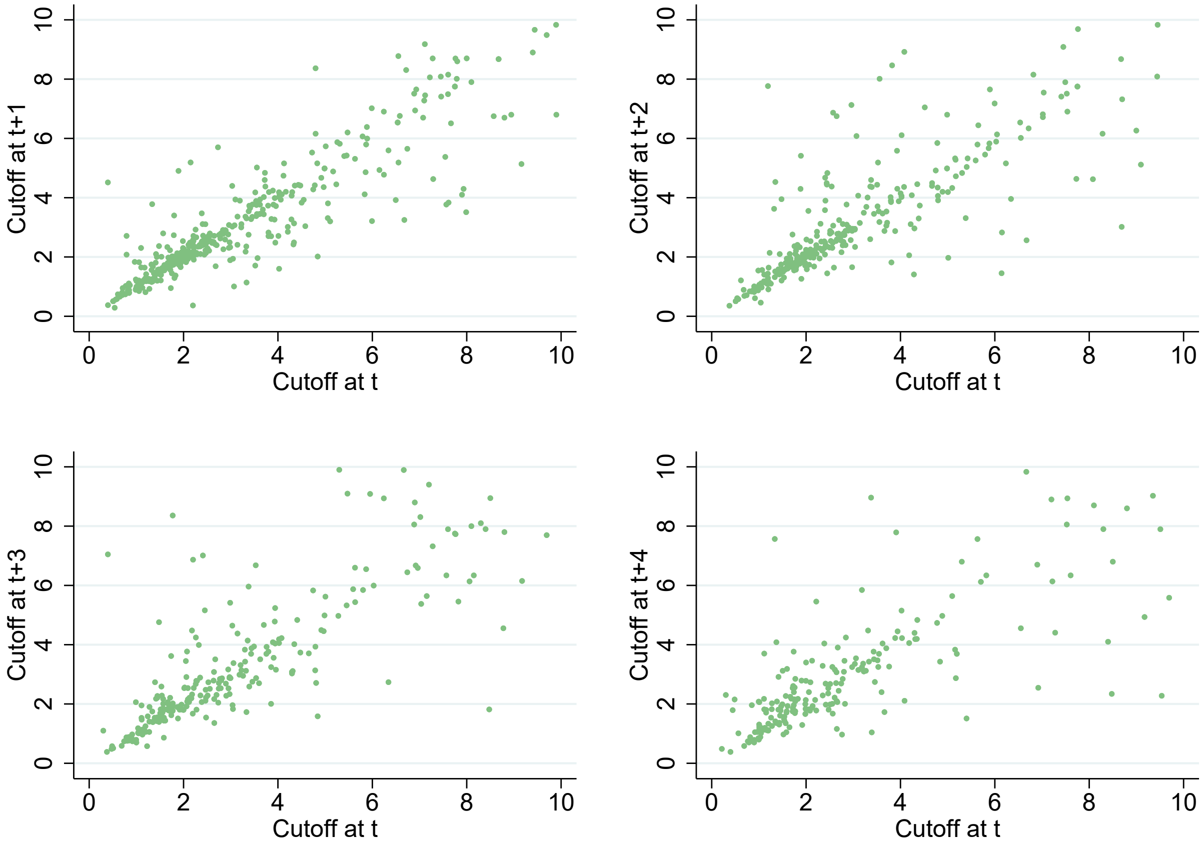
Figure 1: LEAs with brochures



Note: Figure constructed from brochure data, based on whether we any cutoff data within each LEA between 2007 and 2012.

An important premise of our work is that the distance cutoffs are almost impossible to predict, due to the random variation in cohort sizes and applicant preferences. Figure 2 provides supportive evidence of this point by showing the relationship between cutoff distances in year t and in year t+1, t+2, t+3 and t+4 for schools that are observed at least twice in the dataset. Each plot shows just one observation per school. (Even though schools could in fact contribute up to five observations to the top left hand plot of the Figure and up to two observations to the bottom right plot. For each school, we show the observation for which the change in distance between the years in question is the smallest, so as to emphasise the point that these distances bounce around substantially over time). In practice, of the 1,072 points where we have data on cutoffs at t and t+1, just 26 observations (2%) are the same. Just 18 cutoffs are the same at t and t+2, while only 9 are the same at t and t+3. None are the same at longer intervals.

Figure 2: Changes in straight line cutoff distances (in km) over time



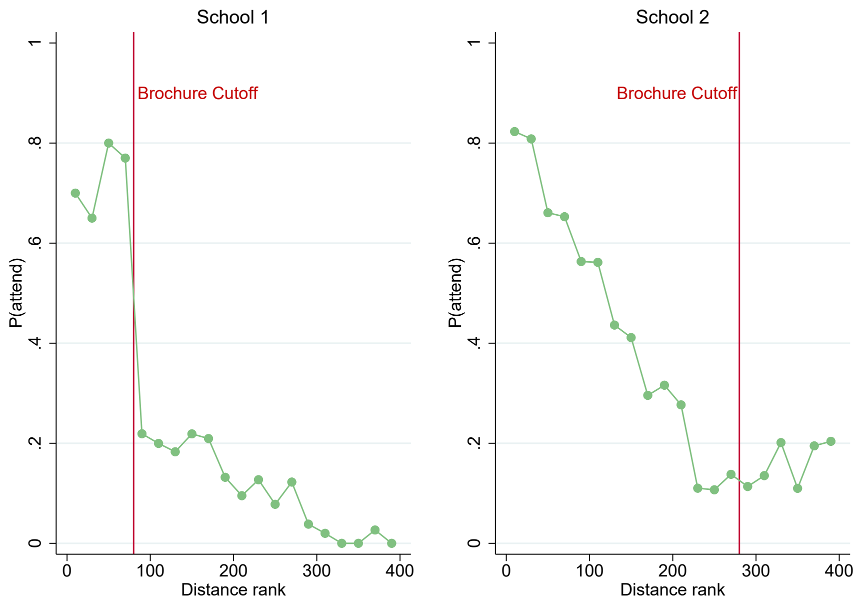
Note: Possible values of t range from 2007 to 2011. To avoid showing more than one observation per school and to emphasise the fact that the cutoffs move around, only the observation that minimises $|y - x|$ by school is shown.

3.3 Sample construction

We define samples at the cutoff-year level. A cutoff-year refers to a school for which there is a known cutoff used in a particular year. We refer to the school as the “focal school". If we observe more than one distance cutoff for a particular school across multiple years, then this school will generate multiple cutoff-year samples. For each cutoff-year a sample consists of all of the students in the relevant year who live in close proximity to the focal school. The "Brochure Sample" is therefore the set of individuals who live near a school with a reported distance cutoff.

Close proximity is based on distance is measured using a student's address measured in January of Year 6. It is measured as straight-line distance from the entrance to the school.[12] We create a new variable, denoted R_{ij}, that describes student i's rank in the distribution of distance to school j. Hence the first-ranked student is the one who lives closest to secondary school j in Year 6, the year before entry, and so on.

Figure 3: Example schools (simulated data)



Note: P(attend) is calculated based on the attendance rate at the school within a bin of n individuals.

In practice, although we have around 2000 brochure cutoffs, many of these will not generate usable attendance discontinuities. An example of this is given in Figure 3, which shows how attendance probabilities change around the brochure cutoff at two example schools (using simulated data). For school 1, the brochure cutoff is associated with a large attendance discontinuity, suggesting the cutoff was used to eliminate a large number of students. For school 2, the brochure cutoff is not associated with a discontinuity in attendance. This is quite common, and happens when the distance criteria was not used to eliminate many students.

[12]In practice, we do not have the precise addresses of students. Instead, we were given distance to the nearest 10 schools (and the identities of those schools).

This analogous to the situation in studies that use lotteries to test for value added bias where there some lotteries which only determine school entry for a very small number of individuals. Abdulkadiroğlu et al. (2020) deal with this by excluding lotteries with at least 100 students who have a nondegenerate risk of attending at least one school.

We generate an equivalent way to select from the full sample of cutoffs to focus only on those which exclude non-trivial numbers of students from attending a school. As we do not have data equivalent to entering a lottery, we instead estimate the following regression model at each of the school-year cutoffs:

$$A_{ij} = \alpha_{1j} + \alpha_{2j}R_{ij} + \alpha_{3j}R_{ij} * Out + \alpha_{4j}Out + \epsilon^A_{ij} \tag{10}$$

where R is individual i's distance rank from school j and Out is a dummy set equal to one if the individual's reported distance from the school is greater than the reported cutoff.[13]

We then exclude schools where there is not a large drop in participation at the brochure cutoff. We also exclude schools where average participation to the left of the cutoff is low. The top panel of Table 3 shows the number of schools-years that would be kept under different selection criteria. When using the reported cutoffs, as in the top panel of the table, we choose selection criteria of a drop in attendance of at least 20% at the cutoff *and* an average participation rate of at least 30% inside the cutoff. This leaves us with 122 school-years, which we refer to as the "Selected Brochure Sample".

We explore whether we can refine the selected brochure sample to account for measurement error in the brochures or in the administrative data. We follow a similar methodology to Hoekstra (2009) by searching for the precise distance cutoff used. Using the brochure cutoff as our starting point, we estimate the equation (10) many times, where each time we vary the definition of "Out" so that we cover every possible rank place within 75 rank places of the reported cutoff. We then choose the definition of "Out" that minimizes the t-statistic on the "Out" coefficient for each school (α_{4j} in equation 10).

The bottom panel of Table 3 shows that this process increases the school-year sample substantially. When we base our data on the estimated cutoffs, we tighten our selection criteria to a drop in attendance of at least 30% at the (estimated) cutoff *and* an average participation rate of at least

[13] The specification includes a linear relationship between rank and attendance that differs either side of the cutoff. We estimate this model including only individuals within 75 rank places of the cutoff for each school.

50% inside the (estimated) cutoff. This leaves us with 267 usable cutoff samples, which we refer to as the "Extended Selected Sample".

Table 3: No. of school-years under different selection criteria

Reported cutoffs					
	Disc > 0% 0	Disc > 10%	Disc > 20%	Disc > 30%	Disc > 40%
P(in)>0	1010	357	137	51	21
P(in)>10%	950	352	137	51	21
P(in)>20%	740	325	130	50	21
P(in)>30%	532	278	*122	48	20
P(in)>40%	373	213	102	44	18
P(in)>50%	263	155	78	36	15
Estimated cutoffs					
P(in)>0	1951	1455	920	443	173
P(in)>10%	1823	1425	914	441	173
P(in)>20%	1401	1234	860	428	171
P(in)>30%	1015	958	745	391	165
P(in)>40%	723	705	589	343	158
P(in)>50%	475	470	426	*267	132

Note: P(in) is the share of students inside the cutoff who attend the focal school. Disc is the attendance discontinuity at the cutoff. * indicates the selection criteria we choose for our main results.

Table 4 shows that almost all of the schools in the brochure sample, and all of the schools in the selected samples are non-selective schools (while 5% of all schools are selective and 15% are special schools). Consistent with Figure 1 above, the brochure schools are more likely to be in London, although a lower share of the selected sample are London based.

Table 4: School level summary stats

	Selected Brochure Sample (1)	Ext. Brochure Sample (2)	Brochure Sample (3)	All schools (4)
School type / location				
Community	36.8%	29.8%	28.9%	23.1%
Academy	52.1%	53.8%	53.0%	53.6%
Selective	0.0%	0.0%	1.1%	4.0%
Special	0.0%	0.0%	0.0%	0.5%
London share	23.1%	24.0%	31.4%	14.2%
N (school-years)	122	267	2262	22579
N (schools)	102	161	770	4573

Note: The all schools sample includes schools appearing up to 6 times in the data to mirror the fact that the same school can have multiple observations in the brochure samples.

3.4 Bias driven by sample selection

A concern with the approach of selecting a subset of school cutoff experiments is that it might in effect involve selecting the experiments which produce a large first stage F statistic. Recent work (Abadie et al., 2019) has critiqued this practice, highlighting that it can generate biased final estimates. However, we argue that our approach is legitimate for three reasons. First, there is a good theoretical case to exclude schools with low participation, and this follows studies which used lotteries to test VAM validity by selecting only the lotteries which randomised a sufficiently large number of students between schools. Second, we show that the correlation between the size of the discontinuity and the discontinuity in value added at the threshold is low (see Table 5), which means that even though there are some cutoffs that generate large value added discontinuities (that is, a large first stage F statistic), these are not necessarily the same schools as the ones with large attendance discontinuities. Third, our specifications pass the basic regression discontinuity checks of whether there are discontinuities in observable characteristics at the cutoff when we use our selected samples (see Section 4.1).

Table 5: Correlation between attendance and value added discontinuities

	Reported cutoffs	Estimated cutoffs
2007	-0.005	-0.055
2008	0.095	0.051
2009	0.055	-0.079
2010	0.081	0.113
2011	0.063	0.094
2012	0.098	0.059

Note: Attendance discontinuities are estimated as in equation 10. Value added discontinuities are estimated in the same way, but replacing attendance with VA as the dependent variable.

3.5 Administrative data

We have administrative data on six cohorts of students who we track through primary and secondary schooling. These cohorts completed primary school and entered secondary school in Autumn of 2007-2012.

Table 6 summarises the pupil-level data, comparing data from all schools to schools with a brochure, to our selected sample of schools from the brochures. The top panel shows a subset of the student background characteristics we include in our models that includes gender, free-school meal (FSM) eligibility, a white British dummy, and an indicator for English not being the students' main language. The panel shows that the brochure school pupils are less white, slightly more likely to qualify for free school meals, and more likely to have English as a second language. This is unsurprising, as schools which use brochure cutoffs are more likely to be in more heavily populated areas, where there are more non white pupils and more poor students. Notably, there is very little difference between the selected brochure samples in columns (1) and (2) and the brochure sample in terms of basic background characteristics. There is almost no discernible difference between columns (3) and (4), which differ because column (3) only includes those who live within 100 rank places of a reported cutoff.

Table 6: Pupil level summary stats

	Selected Brochure Sample (1)	Ext. Brochure Sample (2)	Brochure Sample (3)	Any Brochure Sample (4)	All schools (5)
Student characteristics					
Female	49.5%	52.6%	51.3%	51.3%	49.2%
FSM	15.4%	15.1%	16.9%	17.0%	15.5%
White British	65.0%	64.4%	64.9%	64.6%	77.5%
EAL	28.3%	27.3%	25.4%	25.7%	18.2%
Prior test scores					
Y6 SAT Maths	0.040	0.062	0.036	0.034	0.000
Y6 SAT English	0.027	0.038	0.029	0.029	0.000
Student outcomes					
GCSE scores (capped 8)	0.095	0.127	0.074	0.071	0.000
GCSE scores (all)	0.099	0.141	0.078	0.076	0.000
Longer-run student outcomes (Post 16 education)					
Post GCSE participation	82.2%	81.4%	80.7%	80.5%	74.8%
No. A Levels (if > 0)	2.04	2.04	1.95	1.94	1.90
At least BBB at A Level	0.081	0.087	0.077	0.077	0.073
Observations	24153	52879	426962	680366	3340190
Longer-run student outcomes (Higher education)					
Attended university	0.430	0.448	0.427	0.425	0.366
Attended selective university	0.113	0.121	0.108	0.107	0.094
Observations (HE outcomes)	14610	33210	256305	409450	2262135

Note: Source: ONS. Columns (1)-(3) only include individuals within close proximity to the reported cutoff (within 100 rank places). Column (4) includes everyone who lists a brochure school as one of their nearest 10. Column (5) includes all individuals in the NPD who entered secondary school between 2007 and 2012. GCSE and SAT scores are standardised to have mean zero and a SD for across the whole sample (separately by year to adjust for grade inflation). Observations for higher education ("HE") outcomes are rounded to the nearest 5 following disclosivity rules associated with the HE data. The counts are lower for these as we do not observe HE outcomes for the last two cohorts as they are still too young.

The next panel shows prior attainment information. Children in the brochure sample have slightly better Year 6 scores, but only by 2-3% of a standard deviation. The selected brochure samples generally have higher Year 6 attainment than the brochure schools, especially the extended brochure sample in column (2), but this remains only around 5% of a standard deviation higher than the overall average.

The next panel shows pupil outcomes GCSE outcomes, including best 8 (our primary outcome of interest) and total points scores. GCSE performance is also higher amongst the brochure sample by around 7% of a standard deviation. It is slightly higher again amongst the selected brochure sample (9.5% of a standard deviation higher than the underlying population) and higher still amongst the selected brochure sample (12.7% of a standard deviation higher than the population mean).

Finally, the bottom panel shows longer run student outcomes, including the share staying in education after GCSEs, the number of A-Level (which are typically taken at age 18) passes obtained (amongst those with at least one A Level), the share of students achieving at least BBB at A-Level, the share attending university within two years of finishing school, and the share attending a selective university within two years of finishing school (where selective is defined as being in the top 20% of universities ranked by average entry tariffs). Again, there is higher performance across all these metrics amongst the brochure schools than for all schools, and again the selected sample performance is higher than amongst all brochure schools. To summarize, the table shows that schools in the brochure sample differ from schools in the full sample, in part because they are more likely to be found in densely populated areas. Below, we will focus on the selected and extended brochure samples and compare students either side of the admission cutoffs, as this is ultimately the key test of our research design.

3.6 School effectiveness measures

We test five models of school effectiveness: "Raw scores", which is raw GCSE test score outcomes; "Background" which includes controls for a set of student background characteristics;[14] "Progress 8" which is the main measure of school effectiveness in England; "Lagged scores" which includes controls for Year 6 SAT scores only; and "CVA" which stands for contextualised value added, and which combines background and lagged scores as controls. The final three measures are all considered to be 'value added' measures as they adjust for prior test scores of the students.

[14] These include gender, ethnicity, free-school-meal eligibility and an English as an additional language indicator.

With one exception, these measures are constructed following the regression model given in equation (1), using best 8 GCSE scores as the outcome variable. The exception is Progress 8, which uses a slightly different set of GCSE scores and is estimated using a non-parametric approach.[15] The introduction of Progress 8 was controversial, as previous measures of effectiveness that had been used in England *had* controlled for additional student characteristics Leckie and Goldstein (2017). The government rationalised it by arguing that additional characteristics "*the effect of expecting different levels of progress from different groups of pupils on the basis of their ethnic background, or family circumstances, which we think is wrong in principle*" (DfE, 2010). Our work contributes to this debate by showing whether the inclusion of additional student background characteristics is essential for generating unbiased measures of school effectiveness. The inclusion of the Lagged Scores measure of VA would enable us to see whether any bias in the Progress 8 measure is due to the lack of prior controls or due to the non-parametric approach to estimation.

Table 7: Summary of different school effectiveness measures

			Value added measures		
	Raw scores	Background	Progress 8	Lagged Scores	CVA
Standard deviation	0.847	0.763	0.925	0.438	0.402
Raw scores	1.000				
Background	0.990	1.000			
Progress 8	0.881	0.871	1.000		
Lagged scores	0.903	0.888	0.939	1.000	
CVA	0.888	0.892	0.921	0.980	1.000

Note: Based on data from all schools, pooling the 2007-2012 entry cohorts.

Table 7 shows the standard deviations of each of these effectiveness measures correlation between the raw scores measure of school effectiveness and the value added measures. All of the measures are highly correlated. In general, adjusting for basic background characteristics makes very little difference to the measures of effectiveness, while the effect of including prior test scores is much greater. The correlation of 0.86 between Progress 8 and raw scores is important, as it shows that while there is a strong relationship between raw school performance and Progress 8, there are some schools that perform quite poorly on one measure but well on the other.

[15]The outcome measure, referred to as Attainment 8, doubly weights scores in English and mathematics and then has stricter restrictions on the set of subjects that can be included in the remaining 8 slots. The non-parametric approach involves dividing students into 34 evenly sized bins based on their KS2 SAT results. An individuals' Progress 8 score is then their Attainment 8 score minus the the within-bin average Attainment 8 score (there is also some winsorizing of scores at the bottom end of the distribution). A school's Progress 8 score is the average of the individual Progress 8 scores of each of its pupils.

Table 8 shows how the brochure schools compare to all schools in terms of the different effectiveness measures. The selected schools generally have higher test scores and higher value added scores. This is especially true for the extended brochure sample. Again, we are unconcerned by this as the schools in our sample are from different areas and are oversubsubscribed, so we would not expect them to look the same on average.

Table 8: School level summary stats

	Selected Brochure Sample (1)	Ext. Brochure Sample (2)	Brochure Sample (3)	All schools (4)
Effectiveness measures				
Raw scores	0.146	0.173	0.077	-0.003
Background	0.123	0.143	0.062	-0.002
Progress 8	0.249	0.279	0.186	0.000
Lagged scores	0.088	0.110	0.068	-0.001
CVA	0.067	0.083	0.039	-0.001
N (school-years)	122	267	2262	22579
N (schools)	102	161	770	4573

Note: Based on data from the 2007-2012 entry cohorts.

4 Empirical Framework

Our empirical specification mirrors the approach used with the lottery setting, but with a RD framework. The basic intuition is that we expect the discontinuity in test scores at the cutoffs to align with the discontinuity in value added at the cutoff. That is, if V_j is an unbiased estimator of the value added of school j, we would expect $\pi^Y = \pi^V$ from the following:

$$Y_{ij} = \alpha_1^Y + \alpha_2^Y R_{ij} + \alpha_3^Y (R_{ij} * Out_{ij}) + \pi^Y Out_{ij} + X_i' \Pi^Y + \eta_{ij}^Y \tag{11}$$

$$V_{ij} = \alpha_1^V + \alpha_2^V R_{ij} + \alpha_3^V (R_{ij} * Out_{ij}) + \pi^V Out_{ij} + X_i' \Pi^V + \eta_{ij}^V \tag{12}$$

where R_{ij} is distance rank of individual i from school j and Out_{ij} is a dummy set equal to one for people living just outside the distance cutoff. In practice this would be estimated only for people living near to the cutoff. Including the $(R * Out)$ term allows the relationship between distance rank and the outcome variable to differ inside and outside the cutoff.

As before, this can be recast to a two-stage-least-squares set up with N school-years with usable cutoffs:

$$Y_{ij} = \sum_{j=1}^{N} \theta_j^{MY} C_{ji} + \sum_{j=1}^{N} \delta_j^{MY} C_{ij} R_{ij} + \sum_{j=1}^{N} \gamma_j^{MY} C_{ij} R_{ij} Out_{ij} + \beta^M \hat{V}_{ij} + X_i' \Gamma^{MY} + \zeta_{ij}^{MY} \tag{13}$$

$$V_{ij} = \sum_{j=1}^{N} \theta_j^{MV} C_{ij} + \sum_{j=1}^{N} \delta_j^{MV} C_{ij} R_{ij} + \sum_{j=1}^{N} \gamma_j^{MV} C_{ij} R_{ij} Out_{ij} + \sum_{j=1}^{N} \pi_j^M C_{ij} Out_{ij} + X_i' \Gamma^{MV} + \zeta_{ij}^{MV} \tag{14}$$

Where C_{ij} is a dummy for individual i being in a cutoff sample for focal school j, meaning they live near to the cutoff for that school (this is equivalent to L, the dummy for entering a lottery, from before)[16]. δ_j^{MV} and δ^{MY} capture the slopes for school j inside the cutoff, and γ_j^{MV} and γ_j^{MY} capture the slopes for school j outside the cutoff. As before X_i is a set of student background characteristics. This leaves π_j^M in (14), which is the discontinuity in value added at the cutoff, and captures the effect of our first-stage instruments, and β^M in (13), which is the forecast coefficient. We test whether this equals to one, referring to this as the 'multi-IV' approach (hence the "M" superscript).

Finally, we can also take a 'single-IV' approach, which interacts the instrument, Out, with $\Delta V = V_{ji} - V_{-ji}$, so we have:

$$Y_{ij} = \sum_{j=1}^{N} \theta_j^{SY} C_{ji} + \sum_{j=1}^{N} \delta_j^{SY} C_{ij} R_{ij} + \sum_{j=1}^{N} \gamma_j^{SY} C_{ij} R_{ij} Out_{ij} + \beta^S \hat{V}_{ij} + \tilde{X}_i' \Gamma^{SY} + \zeta_{ij}^{SY} \tag{15}$$

$$V_{ij} = \sum_{j=1}^{N} \theta_j^{SV} C_{ij} + \sum_{j=1}^{N} \delta_j^{SV} C_{ij} R_{ij} + \sum_{j=1}^{N} \gamma_j^{SV} C_{ij} R_{ij} Out_{ij} + \pi^S \sum_{j=1}^{N} C_{ij} Out_{ij} \Delta V_{ij} + \tilde{X}_i' \Gamma^{SV} + \zeta_{ij}^{SV} \tag{16}$$

This is as before but we now estimate a single shift parameter at the cutoff in the V equation, given by π^S, while also including the term V_{-ji} in $\tilde{X}$. Again, our test of bias in V_{ij} comes down to a test of whether β^S, the forecast coefficient, is equal to one. The single-IV approach additionally requires an assumption of V_{-ji}, the school that individual i would have attended had they been outside of the cutoff. For this, we use the most commonly chosen secondary school of primary school peers that is not the focal school.

Rather than estimate a flexible polynomial in the running variable, we follow recommendations from Pei et al. (2022) and estimate a linear slope either side of the threshold but vary the bandwidth used.

[16]In our main specification, we define individual as being in a cutoff sample for a school if they live withing 100 rank places from the cutoff, although we test the sensitivity of this choice in our robustness checks.

4.1 RD specification checks

The validity of our research design depends on the assertion that there is no sorting of parents either side of the brochure cutoff. Previously we argued that this is unlikely given the unpredictability of the admissions process and we showed that the cutoffs bounce around from one year to the next. Here we investigate whether there is evidence of sorting based on observable student characteristics.

We do this by testing for discontinuities in observable student characteristics, $x \subset X$, at the cutoffs, estimated by the following:

$$x_{ij} = \sum_{j=1}^{N} \theta_j^{SX} C_{ij} + \sum_{j=1}^{N} \delta_j^{SX} C_{ij} R_{ij} + \sum_{j=1}^{N} \gamma_j^{SX} C_{ij} R_{ij} Out_{ij} + \pi^{SX} \sum_{j=1}^{N} C_{ij} Out_{ij} + \zeta_{ij}^{SX} \tag{17}$$

Again, these include a full set of school-year dummies and school-year specific slopes either side of the threshold (δ_j and γ_j terms). The results for a range of different characteristics are given for the selected brochure sample and the extended selected sample in Table 9.[17]

Overall, these results strongly favour our research design. The coefficients are all small, and there are very few estimates that are statistically significant at standard levels, suggesting that there are no discontinuous jumps in any of the observable student characteristics at the distance cutoffs. The only counter-example is the 'female' dummy, which is significant at the 50 and 75 bandwidths. However, we think this is most likely driven by sampling noise, as the coefficient is quite noisy at different bandwidths, and also because we think this is the least likely characteristic to change at the threshold.

[17] The top row shows estimates for 'predicted ability', which is the predicted value of test scores, Y_{ij}, from a regression of test scores on the full set of X variables included in our estimation. The remaining rows show a subset of the individual characteristics that are easy to interpret.

Table 9: Testing for discontinuities in student characteristics at the cutoffs

Bandwidth:	25	50	75	100	125
Selected brochure sample (N=122)					
Predicted ability	0.044	0.011	0.000	0.001	-0.003
	(0.036)	(0.025)	(0.020)	(0.018)	(0.016)
p val.	*0.214*	*0.652*	*0.981*	*0.957*	*0.859*
KS2 Maths	0.060	0.044	0.008	0.008	0.000
	(0.050)	(0.035)	(0.028)	(0.024)	(0.022)
p val.	*0.229*	*0.207*	*0.773*	*0.758*	*0.985*
KS2 English	0.001	-0.001	-0.019	-0.005	-0.011
	(0.047)	(0.033)	(0.027)	(0.024)	(0.021)
p val.	*0.979*	*0.972*	*0.492*	*0.831*	*0.608*
FSM	-0.014	0.004	-0.001	0.004	0.004
	(0.019)	(0.014)	(0.011)	(0.010)	(0.009)
p val.	*0.455*	*0.747*	*0.961*	*0.646*	*0.677*
EAL	-0.005	-0.020	-0.014	-0.016	-0.013
	(0.018)	(0.013)	(0.010)	(0.009)	(0.008)
p val.	*0.773*	*0.125*	*0.182*	*0.083*	*0.099*
White British	-0.012	-0.001	-0.008	-0.008	-0.005
	(0.019)	(0.013)	(0.011)	(0.009)	(0.008)
p val.	*0.532*	*0.937*	*0.467*	*0.368*	*0.506*
Female	-0.042	-0.048	-0.031	-0.017	-0.009
	(0.025)	(0.017)	(0.014)	(0.012)	(0.011)
p val.	*0.097*	*0.006*	*0.028*	*0.165*	*0.424*
Obs	6100	12200	18267	24178	29848
Extended selected sample (N=267)					
Predicted ability	-0.002	-0.005	-0.006	-0.017	-0.019
	(0.024)	(0.017)	(0.014)	(0.012)	(0.011)
p val.	*0.948*	*0.770*	*0.645*	*0.166*	*0.086*
KS2 Maths	0.005	-0.012	-0.013	-0.018	-0.023
	(0.033)	(0.023)	(0.019)	(0.016)	(0.015)
p val.	*0.874*	*0.590*	*0.498*	*0.282*	*0.116*
KS2 English	-0.030	-0.012	-0.004	-0.017	-0.018
	(0.033)	(0.023)	(0.019)	(0.016)	(0.015)
p val.	*0.371*	*0.602*	*0.810*	*0.283*	*0.210*
FSM	0.016	0.017	0.012	0.016	0.008
	(0.013)	(0.009)	(0.008)	(0.007)	(0.006)
p val.	*0.217*	*0.068*	*0.111*	*0.015*	*0.164*
EAL	-0.007	-0.003	-0.003	-0.006	-0.006
	(0.012)	(0.009)	(0.007)	(0.006)	(0.006)
p val.	*0.554*	*0.738*	*0.676*	*0.366*	*0.277*
White British	0.010	-0.002	-0.006	-0.006	-0.002
	(0.013)	(0.009)	(0.007)	(0.007)	(0.006)
p val.	*0.430*	*0.803*	*0.399*	*0.371*	*0.800*
Female	-0.003	0.000	0.001	-0.001	-0.004
	(0.017)	(0.012)	(0.009)	(0.008)	(0.007)
p val.	*0.847*	*0.967*	*0.942*	*0.910*	*0.610*
Obs	13350	26692	39919	52853	65291

Note: RD bandwidth is based on rank places either side of the cutoff. Standard errors are given in the parentheses.

In Appendix Table A1, we show results when we interact the Out dummy with ΔV (and include an additional control for $_{-ji}$, following the single-IV setup above. For the selected sample, the results are very similar, but for the extended selected sample they are less convincing. While the background characteristics (free school meal eligibility, English as an additional language, white British) are all pass the tests, there are several estimates, including predicted ability as well as prior maths and English scores that do not. While this is not the case at all bandwidths, our takeaway from these results is that we should be cautious in using the searching method to identify cutoffs. It seems that even a relatively tight search around the reported cutoff can pick up natural barriers around which parents select. Based on this, although we have to work with a smaller sample when we focus on the precise brochure cutoffs, we prefer that as to the potentially bias-inducing approach of searching for cutoffs.

5 Results

5.1 Main estimates

Our main estimates of the forecast coefficient are shown Table 10. These are our main estimates as they test for bias in measures of secondary school effectiveness based on test scores taken at the end of secondary school (GCSEs). The different columns show estimates of the forecast coefficient for the five sets of effectiveness measures introduced in Section 3.6. In all cases we cluster standard errors at the primary school level.

We show estimates for the single-IV and for multi-IV approaches using 2SLS and LIML. All three approaches find that the effectiveness measures which do not adjust for prior test scores are biased. In both columns (1) and (2), the forecast coefficients are around 0.5, and we strongly reject the null that the coefficient is equal to one.

Table 10: Forecast coefficient test, selected brochure sample

			Value added measures		
	Raw scores	Background	Progress 8	Lagged Scores	CVA
	(1)	(2)	(3)	(4)	(5)
Single IV					
Forecast coefficient	0.515	0.563	0.956	0.897	0.896
s.e.	(0.096)	(0.105)	(0.133)	(0.157)	(0.167)
p value	*0.000*	*0.000*	*0.739*	*0.513*	*0.532*
First Stage F Stat.	149.9	144.0	181.3	180.4	172.7
N		Schools: 102	School-years: 122		Pupils: 23,766
Multi IV (2SLS)					
Forecast coefficient	0.503	0.576	0.837	0.933	1.002
s.e.	(0.109)	(0.117)	(0.143)	(0.157)	(0.162)
p value	*0.000*	*0.000*	*0.256*	*0.670*	*0.990*
First Stage F Stat.	2.3	2.4	3.0	3.4	3.5
OverID test p value	*0.078*	*0.100*	*0.053*	*0.232*	*0.266*
N		Schools: 102	School-years: 122		Pupils: 24,024
Multi IV (LIML)					
Forecast coefficient	0.546	0.622	0.801	0.950	1.008
s.e.	(0.156)	(0.161)	(0.183)	(0.190)	(0.194)
p value	*0.004*	*0.019*	*0.278*	*0.792*	*0.967*
First Stage F Stat.	2.3	2.4	3.0	3.4	3.5
OverID test p value	*0.072*	*0.093*	*0.048*	*0.219*	*0.252*
N		Schools: 102	School-years: 122		Pupils: 24,024

Note: Outcome is best 8 GCSE point score for columns 1,2, 4 and 5 and Attainment 8 for column 3.

However, we are unable to reject the null that any of our value added estimates. The single IV estimates are all 0.9-0.95, with p values of 0.5 or above. For the multi-IV approaches, we get forecast coefficient estimates of close to 1 for the lagged score and CVA approaches, with p values of around 0.7 or above. Also reported for the multi-IV models is the p-value from the overidentification test proposed by Angrist et al. (2017). This essentially measures whether the estimator has the same predictive power in every school-year cutoff experiment. The results in columns (4) and (5) show that we also fail to reject the null for this test for the lagged scores and CVA estimates.

The multi-IV tests of Progress 8 bias are a little less compelling than the results in columns (4) and (5). Although we still do not reject the null of unbiasedness, the coefficients are closer to 0.8 and we borderline reject the null for the overidentification test at the 5% level. Overall, these results suggest that VAMs that control *only* for prior attainment are sufficient to alleviate bias. This is more marginal in the case of Progress 8, which might suggest that Progress 8 does not control for lagged scores in a flexible enough way, but the results do not provide a compelling case that the measure is systematically biased.[18]

A concern about the multi-IV estimates is the low F-statistic, which is indicative of a weak IV problem. This is likely driven by schools not ultimately being that different either side of the cutoff, meaning that just being outside the threshold does not shift value added by very much. This could be problematic in the case of the forecast test, as it would mean the IV estimate is biased towards the OLS estimate of effectiveness. This is why we also provide LIML estimates of the multi-IV case, as this estimator performs well in the case of weak instruments (Angrist and Frandsen, 2022). The results using LIML are very similar to the 2SLS case, which we find to be reassuring.

5.2 Robustness

Table 11 shows our estimates of the forecast coefficients for our five sets of effectiveness measures under a range of different robustness checks. In the first panel, we show that our results are the same when we use the extended brochure sample, which is reassuring despite there being question marks around the validity of this approach.

Next, we show that the precise selection criteria we use from our brochure sample do not drive our results. In the second panel we tighten them so that the participation rate within the cutoff has to be higher (at least 40%) and the attendance discontinuity has to be bigger (at least 30%), relative to our choices of 30% and 20% in the baseline case. This reduces our number of school-year experiments to just 102 (see Table 3), but our conclusions are unchanged - that is, the measures of school effectiveness in columns (1) and (2) are rejected, while the VAMs that adjust for prior attainment are found to be unbiased. Again, the inclusion of background characteristics in addition to prior attainment adds little to the precision of the estimates.

[18]Notably, a 2018 amendment to Progress 8 involved the introduction of a floor on scores to prevent very negative outliers from skewing the results. When we estimate Progress 8 without this amendment, we do find systematic evidence that the measure is biased.

Table 11: Forecast coefficient test, robustness checks

	Raw scores	Background	Value added measures: Progress 8	Lagged Scores	CVA
	(1)	(2)	(3)	(4)	(5)
Extended Selected Sample					
Forecast coefficient	0.479	0.551	0.917	0.848	0.891
s.e.	(0.068)	(0.077)	(0.148)	(0.127)	(0.137)
p value	*0.000*	*0.000*	*0.576*	*0.230*	*0.429*
N		Schools: 161	School-years: 267		Pupils: 52,212
Tighter selection (P(in)>40%, Disc > 20%)					
Forecast coefficient	0.483	0.534	0.865	0.835	0.869
s.e.	(0.097)	(0.107)	(0.136)	(0.160)	(0.169)
p value	*0.000*	*0.000*	*0.322*	*0.301*	*0.439*
N		Schools: 85	School-years: 102		Pupils: 19,822
Looser selection (P(in)>20%, Disc > 10%)					
Forecast coefficient	0.459	0.509	0.730	0.699	0.719
s.e.	(0.103)	(0.114)	(0.144)	(0.177)	(0.190)
p value	*0.000*	*0.000*	*0.061*	*0.089*	*0.138*
N		Schools: 226	School-years: 235		Pupils: 63,875
Bandwidth = 75 (Baseline 100)					
Forecast coefficient	0.509	0.561	0.954	0.915	0.904
s.e.	(0.118)	(0.130)	(0.165)	(0.190)	(0.201)
p value	*0.000*	*0.001*	*0.781*	*0.656*	*0.632*
N		Schools: 102	School-years: 122		Pupils: 17,933
Bandwidth = 125					
Forecast coefficient	0.489	0.546	0.905	0.888	0.917
s.e.	(0.085)	(0.093)	(0.122)	(0.146)	(0.157)
p value	*0.000*	*0.000*	*0.435*	*0.442*	*0.598*
N		Schools: 102	School-years: 122		Pupils: 29,388
Outcome: total GCSE points					
Forecast coefficient	0.579	0.634	-	0.954	0.965
s.e.	(0.090)	(0.096)	-	(0.135)	(0.138)
p value	*0.000*	*0.000*	-	*0.731*	*0.802*
N		Schools: 102	School-years: 122		Pupils: 23,766

Note: Robustness checks are shown for the single-IV estimates.

In the third panel we loosen the selection criteria, so that the participation rate within the cutoff only needs to be at least 20%, while the drop in attendance at the cutoff is only required to be 10%. This reduces the forecast coefficient to around 0.7 and it becomes only a borderline pass of the test.

The next two panels show that our main results are not sensitive to the RD bandwidth used. This check is essential as it shows that our findings are not driven by our choice to impose the restriction of linear (school-specific) slopes either side of the threshold. Combining the fact that we allow these slopes to differ either side of the threshold and the robustness to different bandwidths suggests that the findings are not sensitive to our choice of specification.

Finally, the table shows the sensitivity to the choice of outcome variable, using total GCSE points, rather than points from each student's best eight GCSEs (including English and maths). The overall results are again very similar.

5.3 Heterogeneous treatment effects and parental inputs

One concern with the forecast coefficient test is that it assumes homogeneous treatment effects of schools. As pointed out by Angrist et al. (2017), rejection of the null hypothesis could be driven by heterogeneous treatment effects, rather than by unobserved selection bias. In our context, it could be the case that pupils who live close to these cutoffs might be systematically different and might therefore experience a different effect from getting into the focal school than the average student. Ultimately, we consider this issue to be less important because we do not reject the null that our value added measures are unbiased. The failure to reject the null for the overidentification test - at least in the case of the lagged scores and CVA models - is also evidence to support our defence that our findings are driven by this issue.

We make a similar argument about the role of parental inputs: one concern might be that parents who just miss out on their preferred school compensate by increasing inputs (Greaves et al., 2023, show this mechanism to be important in other contexts). Ultimately, this effect would bias down our estimate of the forecast coefficient,[19] which makes an invalid rejection more likely. Since we do not have measures of parental inputs, this is a difficult hypothesis to test. However, again the fact that we do not reject the null for our value added measures us less concerned about this issue.

5.4 Longer run impacts

We also consider whether our effectiveness measures are predictive of longer run outcomes in Table 12. For this exercise, we replace the outcome variable in the second stage equation (see equation 15) with longer run outcomes, including the number of A-Level grades obtained, whether or not student achieve at least three B grades in their A-Level exams, whether students attend university by age 19, and whether they attend a selective university by age 19.

[19] As described above, the forecast coefficient is equal to the ratio of the discontinuity in test scores to the discontinuity in value added, $\beta = \alpha_3^Y / \alpha_3^V$ from equations (11) and (12) above. This mechanism would reduce α_3^Y, and hence bias down the estimate of β.

Table 12: Longer run outcomes

	Raw scores (1)	Background (2)	Value added measures: Lagged Scores (3)	Value added measures: CVA (4)
Total A-Levels				
Forecast coefficient	0.597	0.625	0.759	0.759
s.e.	(0.128)	(0.136)	(0.179)	(0.185)
p value	*0.002*	*0.006*	*0.179*	*0.193*
N	Schools: 102	School-years: 122	Pupils: 23,890	
Achieve at least BBB at A-Level				
Forecast coefficient	0.742	0.720	1.165	1.024
s.e.	(0.283)	(0.296)	(0.496)	(0.516)
p value	*0.362*	*0.344*	*0.740*	*0.962*
N	Schools: 102	School-years: 122	Pupils: 23,935	
Attend university				
Forecast coefficient	0.600	0.644	0.771	0.907
s.e.	(0.274)	(0.292)	(0.452)	(0.478)
p value	*0.145*	*0.222*	*0.613*	*0.846*
N	Schools: 65	School-years: 74	Pupils: 14,405	
Attend selective university				
Forecast coefficient	0.669	0.687	1.231	1.197
s.e.	(0.324)	(0.342)	(0.530)	(0.552)
p value	*0.307*	*0.359*	*0.663*	*0.722*
N	Schools: 65	School-years: 74	Pupils: 14,405	

Note: Source: ONS. Longer run outcomes are shown for the single-IV estimates. Sample sizes are rounded to the nearest 5 in accordance with disclosivity rules.

This exercise is potentially interesting, as it is possible that parents select into school based on factors that influence longer-term outcomes of children. Therefore, it is plausible that the short run effectiveness would be unbiased but they longer run measures would not. However, unfortunately our test appears to lack adequate power to be able to say anything informative on this issue (the Table shows that our standard errors are generally much larger).

6 Do parents leave value added on the table?

We now draw on the result that we have unbiased estimates of secondary school effectiveness to investigate whether there are potential gains to be made from interventions in application behaviours of parents. To do this, we make use of additional data we have on school applications made by parents in 2014.

Table 13: Potential achievement gains from ranking schools based on effectiveness

	Average value added			SES Gap
	Overall	High SES	Low SES	
	(1)	(2)	(3)	(4)
Observed first preference	0.212	0.549	-0.109	0.658
	(55.9)	(67.4)	(44.9)	(22.5)
Nearest school	0.044	0.327	-0.217	0.545
	(50.2)	(60.2)	(41.2)	(18.9)
Choosing the highest VA school, given:				
$dist_n \leq dist_o$	0.407	0.682	0.168	0.514
	(62.2)	(71.5)	(54.4)	(17.1)
$dist_n \leq dist_o + 1km$	0.722	0.875	0.668	0.207
	(72.4)	(77.3)	(70.9)	(6.4)
$dist_n \leq 1km$	0.275	0.518	0.112	0.406
	(58.0)	(66.5)	(52.5)	(14.0)
$dist_n \leq 3km$	0.849	0.915	0.911	0.003
	(76.7)	(78.6)	(79.0)	(-0.4)
N	409,112	95,184	73,017	

Note: Value added estimates adjust for prior attainment and background scores and are standardised to have a mean of zero with a standard deviation of 1 in the baseline population. Percentile ranks are given in the parentheses. High SES includes individuals in the top quintile for SES while Low SES is the bottom quintile (SES is based on the IDACI index, a measure of income deprivation). $dist_n$ is distance to the newly assigned first choice school and $dist_o$ is observed distance to current school. We exclude individuals without preference data and without an observed value for $dist_o$ which cuts the sample size by about 20%.

Table 13 shows the average value added (in standard deviations, σ) of parents' first choice schools under different selection criteria. The first row shows the average value added of the observed first preference school, overall and split by student socio economic status (SES). The figures show that on average, the first preference school is 0.21σ above the overall average, or at the 56th percentile of all schools. It also shows that there is a large SES gap in first choice school value added of 0.66σ (around 23 percentiles), as parents from high SES backgrounds apply for schools that are 0.55σ above average, while parents from low SES backgrounds apply for schools that are slightly below (0.11σ) below average.

The second row shows the average value added of the *nearest* school to each individual.[20] Notably, the SES gap in value added is considerably smaller (0.54σ, or 19 percentiles) than in the first row, suggesting that higher SES parents are more willing or able to travel further to attend more effective schools.

The following set of four rows of results then show how these numbers would be affected

[20] This excludes single (opposite) sex schools and independent schools.

if parents were to apply to their most effective local school, where each of the four rows has a different definition of "local". In the first, we constrain it so that parents cannot select a school that is further away than they one their child actually attended. In the second row, we allow them to select a school that is up to 1km further away from their home than the school their child actually attended. We then make more straightforward restrictions of selecting the most effective school within a 1km radius of their home, and finally within 3km of their home.[21]

The figures from column (1) suggest students would lose out considerably by enrolling in their first choice school rather than the most effective school locally. This loss is around a 20% of a standard deviation on average (the difference between 0.407, the value added of the most effective school within the radius of the individuals' actual school, and 0.212, the average value added of the first preference school) when we constrain travel distance to be no further than the school actually attended. It is much larger (around 50% of a standard deviation) if we extend the local search radius to the actual distance plus 1km. It is 65% of a standard deviation if parents choose the most effective school within 3km of their home (and 6% of a standard deviation within 1km of their home). These are very large potential gains in GCSE scores. These results are consistent with Ainsworth et al. (2022) in that they suggest parents are leaving value added "on the table" when choosing a secondary school for their children.[22]

The remaining columns suggest that this conclusion is more pronounced for those from lower SES backgrounds. The gap in value added of first choice schools of 0.66σ would drop to 0.51σ if distance is constrained to being no further than the distance to the actual school attended. It also drops considerably in the other scenarios - most notably, the SES gap declines to effectively zero if parents were to select the most effective school within 3km of their home.

7 Conclusion and discussion

This paper tests for bias in VAMs in a setting where school admissions lotteries do not exist. Instead we exploit discontinuities in attendance probabilities that occur due to distance being used as a tiebreaker in school admissions processes in England. While previous studies have

[21] In our 2014 data, around 50% of pupils in the state system attend a school within 1km of their home, while 95% attend a school within 3km.

[22] In theory, a school application process would be "strategy proof" in that it would allow parents to apply for as many schools as they like, ordering their preferences truthfully, as the centralised algorithm does not allow schools to select on preference ranking. However, because in practice many parents are only allowed to select three options, they are likely to be strategic in their application behaviour, as they might not want to risk not being allocated to any of the schools that they apply to (Walker and Weldon, 2020).

used school entry lotteries in the United States to test for bias in VAMs commonly used there, the evidence is very mixed and in any case may well not apply to the English setting.

This issue is important within the institutional context due to the use of 'Progress 8' as the primary measure of school effectiveness. This measure *only* adjusts school outputs for prior attainment, which has generated some criticism by academics. We show that controlling for prior attainment is essential to generate unbiased estimates of school effectiveness - but also that measures that only control for prior attainment in England are sufficient to capture true effectiveness.

The result is also important more generally, as it contributes to the small body of evidence that suggests very simple controls for prior attainment are sufficient to generate unbiased estimates of effectiveness (Deming, 2014). This is useful because such measures are much less demanding in terms of data requirements, and they are also easier for schools to understand and target. Indeed the latter factor might provide an explanation for why Progress 8 has become such a prominent feature of the English system, in contrast to previous measures of value added (such as the 'contextualised value added' measure used in the mid-2000s).

Finally, we show that considerable test score improvements could be made if parents were to select school based on their effectiveness. The results, which are consistent with Ainsworth et al. (2022) and Abdulkadiroğlu et al. (2020), suggest that parents are leaving value added "on the table" when choosing secondary schools. This could be because they see other school characteristics as more important (for example Burgess et al., 2015, show that distance is a highly important determinent of preferences), or because they are unaware of effectiveness measures or fail to understand them. This finding suggests that information campaigns to increase awareness and the importance of attending more effective schools could potentially result in quite large overall gains and could also narrow gaps between students from richer and poorer backgrounds.

References

Abadie, Alberto, Jiaying Gu, and Shu Shen, "Instrumental variable estimation with first stage heterogeneity," Technical Report, Tech. rep 2019.

Abdulkadiroğlu, Atila, Parag A Pathak, Jonathan Schellenberg, and Christopher R Walters, "Do parents value school effectiveness?," *American Economic Review*, 2020, *110* (5), 1502–39.

Ainsworth, Robert, Rajeev Dehejia, Cristian Pop-Eleches, and Miguel Urquiola, "Why do households leave school value added "on the table"? The roles of information and preferences,"," *American Economic Review*, 2022.

Andrabi, Tahir, Natalie Bau, Jishnu Das, and Asim Ijaz Khwaja, "Heterogeneity in School Value-Added and the Private Premium," Technical Report, National Bureau of Economic Research 2022.

Angrist, Joshua D and Brigham Frandsen, "Machine labor," *Journal of Labor Economics*, 2022, *40* (S1), S97–S140.

Angrist, Joshua D., Peter D. Hull, Parag A. Pathak, and Christopher R. Walters, "Leveraging Lotteries for School Value-Added: Testing and Estimation*," *The Quarterly Journal of Economics*, May 2017, *132* (2), 871–919.

Angrist, Joshua, Peter Hull, and Christopher R Walters, "Methods for Measuring School Effectiveness," 2022.

_ , _ , Parag A Pathak, and Christopher Walters, "Credible school value-added with undersubscribed school lotteries," *The Review of Economics and Statistics*, 2021, pp. 1–46.

Bau, Natalie and Jishnu Das, "Teacher value added in a low-income country," *American Economic Journal: Economic Policy*, 2020, *12* (1), 62–96.

Burgess, Simon, Ellen Greaves, Anna Vignoles, and Deborah Wilson, "What parents want: School preferences and school choice," *The Economic Journal*, 2015, *125* (587), 1262–1289.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff, "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, September 2014, *104* (9), 2593–2632.

Deming, David J., "Using School Choice Lotteries to Test Measures of School Effectiveness," *American Economic Review, Papers and Proceedings*, May 2014, *104* (5), 406–411.

DfE, "The Importance of Teaching: The Schools White Paper 2010," *Dept. for Education*, 2010.

Greaves, Ellen, Iftikhar Hussain, Birgitta Rabe, and Imran Rasul, "Parental responses to information about school quality: Evidence from linked survey and administrative data," *Economic Journal*, 2023.

Hoekstra, Mark, "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach," *The Review of Economics and Statistics*, October 2009, *91* (4), 717–724.

Kane, Thomas J and Douglas O Staiger, "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation," Working Paper 14607, National Bureau of Economic Research December 2008. Series: Working Paper Series.

_ , Daniel F McCaffrey, Trey Miller, and Douglas O Staiger, "Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment. Research Paper. MET Project.," *Bill & Melinda Gates Foundation*, 2013.

Leckie, George and Harvey Goldstein, "The evolution of school league tables in England 1992–2016: 'Contextual value-added', 'expected progress' and 'progress 8'," *British Educational Research Journal*, 2017, *43* (2), 193–212.

Pei, Zhuan, David S Lee, David Card, and Andrea Weber, "Local polynomial order in regression discontinuity designs," *Journal of Business & Economic Statistics*, 2022, *40* (3), 1259–1267.

Walker, Ian and Matthew Weldon, "School choice, admission, and equity of access: Comparing the relative access to good schools in England," 2020.

Appendix

A Cutoff data

Figure A1: Example brochure

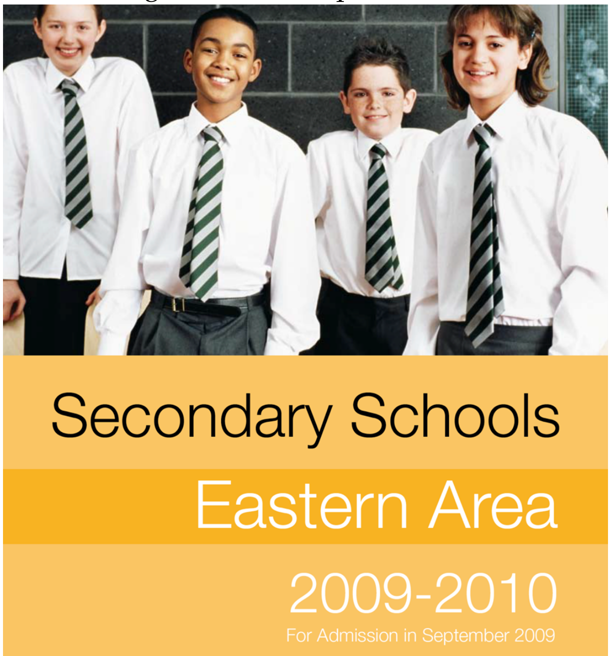


Figure A2: Example inside the brochure

School Name, Headteacher, Address, Telephone, Type and Age Range	Admission Number & Number of Applicants
Kenilworth School & Sports College Mr H. Abbott Leyes Lane, Kenilworth, CV8 2DA Tel: 01926 859421 Fax: 01926 859426 Email: Kenilworth.school@ksn.org.uk Website: www.ksn.org.uk Community, 11-18 Mixed Comprehensive	270 432 Total preferences of which 277 were first preferences Offers made to all on time ctieria 1,2,3,4 & 5 to a distance of 1.861 miles

Kenilworth School

Name of school	Type of school	Places available	Cut off distance as at 3 March 2008
Ashmole	Mixed	224	0.670 miles
Hendon	Mixed	200	N/A
Mill Hill High	Mixed	240	0.863 miles
Queen Elizabeth's	Boys Selective	180	N/A

Barnet Schools

B Additional RD checks

Table A1: Testing for discontinuities in student characteristics at the cutoffs

Bandwidth:	25	50	75	100	125
Selected brochure sample (No. schools=122)					
Predicted ability	0.117	-0.028	-0.057	-0.068	-0.067
	(0.110)	(0.074)	(0.060)	(0.052)	(0.046)
p val.	*0.287*	*0.704*	*0.340*	*0.189*	*0.148*
KS2 Maths	0.148	-0.116	-0.126	-0.099	-0.068
	(0.151)	(0.097)	(0.079)	(0.068)	(0.061)
p val.	*0.326*	*0.234*	*0.112*	*0.147*	*0.266*
KS2 English	0.121	-0.069	-0.130	-0.094	-0.106
	(0.142)	(0.097)	(0.079)	(0.069)	(0.061)
p val.	*0.395*	*0.474*	*0.101*	*0.169*	*0.081*
FSM	-0.130	-0.072	-0.061	-0.040	-0.037
	(0.061)	(0.043)	(0.034)	(0.029)	(0.026)
p val.	*0.033*	*0.093*	*0.075*	*0.167*	*0.161*
EAL	-0.009	0.040	0.041	0.030	0.020
	(0.057)	(0.039)	(0.032)	(0.028)	(0.024)
p val.	*0.877*	*0.316*	*0.203*	*0.272*	*0.402*
White British	0.074	0.004	-0.013	0.000	0.007
	(0.055)	(0.037)	(0.030)	(0.026)	(0.023)
p val.	*0.183*	*0.916*	*0.665*	*0.994*	*0.751*
Female	-0.030	-0.068	-0.047	-0.026	-0.019
	(0.075)	(0.051)	(0.042)	(0.036)	(0.032)
p val.	*0.688*	*0.178*	*0.255*	*0.471*	*0.555*
Obs	6100	12200	18267	24178	29848
Extended selected sample (No. schools=267)					
Predicted ability	-0.060	-0.077	-0.085	-0.112	-0.115
	(0.068)	(0.047)	(0.038)	(0.033)	(0.030)
p val.	*0.381*	*0.100*	*0.024*	*0.001*	*0.000*
KS2 Maths	-0.002	-0.093	-0.085	-0.105	-0.110
	(0.091)	(0.062)	(0.050)	(0.043)	(0.039)
p val.	*0.979*	*0.136*	*0.092*	*0.016*	*0.005*
KS2 English	-0.068	-0.093	-0.072	-0.106	-0.092
	(0.092)	(0.063)	(0.051)	(0.044)	(0.039)
p val.	*0.461*	*0.142*	*0.156*	*0.015*	*0.020*
FSM	0.032	0.007	0.020	0.018	0.032
	(0.039)	(0.027)	(0.022)	(0.019)	(0.017)
p val.	*0.413*	*0.790*	*0.357*	*0.336*	*0.061*
EAL	-0.058	-0.029	-0.034	-0.034	-0.021
	(0.034)	(0.024)	(0.020)	(0.017)	(0.015)
p val.	*0.087*	*0.218*	*0.083*	*0.044*	*0.152*
White British	0.060	0.042	0.033	0.016	0.013
	(0.035)	(0.024)	(0.020)	(0.017)	(0.015)
p val.	*0.086*	*0.084*	*0.098*	*0.352*	*0.385*
Female	0.020	-0.012	-0.043	-0.050	-0.045
	(0.043)	(0.030)	(0.025)	(0.022)	(0.020)
p val.	*0.650*	*0.692*	*0.086*	*0.020*	*0.022*
Obs	13350	26692	39919	52853	65291

Note: RD bandwidth is based on rank places either side of the cutoff. p values are given in the parentheses. N is the number of school-year experiments.