THE ECONOMIC JOURNAL

The Economic Journal, **125** (*September*), 1262–1289. Doi: 10.1111/ecoj.12153 © 2014 The Authors. The Economic Journal published by John Wiley & Sons Ltd on behalf of Royal Economic Society. Published by John Wiley & Sons, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

WHAT PARENTS WANT: SCHOOL PREFERENCES AND SCHOOL CHOICE*

Simon Burgess, Ellen Greaves, Anna Vignoles and Deborah Wilson

We investigate parents' preferences for school attributes in a unique data set of survey, administrative, census and spatial data. Using a conditional logit, incorporating characteristics of households, schools and home–school distance, we show that most families have strong preferences for schools' academic performance. Parents also value schools' socio-economic composition and distance, which may limit the potential of school choice to improve academic standards. Most of the variation in preferences for school quality across socio-economic groups arises from differences in the quality of accessible schools rather than differences in parents' preferences, although more advantaged parents have stronger preferences for academic performance.

Strong parental demand for academic performance is a central element of the view that strengthening school choice will drive up school performance (Hastings *et al.*, 2008). As school choice is a widely endorsed school improvement policy, this assumption is also an important policy issue, and the academic and policy debates on school choice are both controversial and unresolved (Hoxby, 2003). We contribute to this debate by offering new evidence on the nature and heterogeneity of parents' preferences for schools. We address three key questions. First, what school attributes do families value? Is the school's academic attainment record important, or do other factors out-weigh it? Second, how much do preferences differ between families of different socio-economic status? Answering these questions helps to explain the disproportionate admission of children from poor families to academically low-performing schools (Burgess and Briggs, 2009). Finally, we provide evidence on the degree to which this arises through differences in preferences for school attributes as opposed to constraints caused by differences in the attributes of available accessible schools.

To address these questions we assemble a unique data set. We use survey information on parents' primary school choices and a rich set of family socio-economic and neighbourhood characteristics. We link this to administrative data on the characteristics of schools, and the nature of the local school choice mechanism. To identify

* Corresponding author: Simon Burgess, Centre for Market Priory Road, Bristol, BS8 1TX, UK. Email: simon.burgess@bristol.ac.uk.

Many thanks to Jon Johnson, Senior Database Manager at the Centre for Longitudinal Studies for providing access and guidance for using secure data from the Millennium Cohort Study, and to Kirstine Hansen, Research Director of the Millennium Cohort Study for her support. Thanks also to DCSF for access to the NPD data set. We thank the Editor and referees for very helpful comments. Many thanks also to Rebecca Allen, Estelle Cantillon, Steve Gibbons, Justine Hastings, Caroline Hoxby, Parag Pathak, Patricia Rice, Frank Windmeijer and seminar participants at NBER, Oxford, Warwick, Ca'Foscari, Leuven and Bristol for very helpful comments. Finally, we thank the ESRC for funding this project (RES-000-22-2843).

Correction Note: This article was first published online on the 21st of August 2014, under a subscription publication licence. The article has since been made OnlineOpen, and the copyright line and licence statement was therefore updated in April 2015.

parents' constraints in terms of available school choices, we use the national pupil census with embedded spatial information to model *de facto* catchment areas around schools within which there is a high probability of admission. The strengths of our data complement the data set used by Hastings *et al.* (2008) in their analysis of parents' preferences.

'School choice' in England means the right to express a preference for particular schools and for each parent's highest possible preference to be honoured (according to a pre-defined allocation mechanism). Parents can nominate between three and six choices, in order of preference, on the appropriate school district (local authority (LA)) form. The components of the decision are as follows: the set of schools that the parents are choosing from; information on the attributes of those schools; parents' preferences over those characteristics and an allocation mechanism that maps from the parents' nomination to actual school attended.

We face two central identification challenges. First, we must define the *de facto* set of schools that a family can actually choose from, as there are no legal or natural restrictions on this in England.[1] We combine the pupil census with spatial information to define a minimal set of schools with a high probability of admission (conditional on application) for each household in the survey. We assess the robustness of our results to a different definition of the feasible choice set (FCS), which includes all schools within a certain distance from the home and does not account for the probability of admission to each school.

The second identification challenge also relies on the definition of the FCS of schools for each household: determining how the school nominated by the family relates to their true preferences as a function of the allocation mechanism. Parents' nominations are likely to be affected by the admission criteria (or the mechanism design) used by the LA. This issue is informed by the mechanism design literature that has analysed agents' optimal responses to assignment mechanisms (Abdulkadiroğlu and Sönmez, 2003; Abdulkadiroğlu *et al.*, 2009). We believe that the school choices we observe are likely to reflect true preferences for school attributes, as at the relevant date for our survey about two thirds of LAs in England used a mechanism likely to elicit truthful preferences among schools that have a high probability of admission (Coldron *et al.*, 2008). To explore the validity of this assumption, one of the robustness checks restricts the sample to those areas where the allocation mechanism is truth revealing.

To address our first and second questions on parents' preferences for school attributes and the variation in observed preferences between socio-economic groups, we run a conditional logit model on households' top school choice. We show that families care about three main school attributes: the academic quality of the school, its socio-economic composition and the home–school distance. The majority of households prefer schools with higher academic standards. On average, families prefer schools with fewer children living in low-income households. Almost all households have strong preferences for proximity. Preferences appear to be heterogeneous across socio-economic (SES) groups: those in the lowest SES group in particular have distinct preferences, with negative demand responses to increases in academic quality and

[1] For instance, families can choose schools outside their school district (LA).

positive demand responses to decreases in the socio-economic composition of the school. Households from each SES group value proximity to the same extent, however, suggesting that parents in each group are equally willing to travel to a school that is feasible and meets their other preferences. We are mindful, however, that this model may still to a degree conflate household preferences with the constraints they face.

To provide evidence on the relative roles of preferences and constraints in driving differences in chosen school quality between SES groups (our third key question) we adopt a different approach. We first show that there are big differences in the attributes of accessible schools between households of high and low socio-economic status. Because of the spatial clustering of our survey data, the majority of our sample share feasible school choice sets with at least one other household. This allows us to define choice set fixed effects and therefore to compare the choices of households confronted with exactly the same choices. Estimating the model with and without the fixed effects means that we can decompose the overall relationship between SES and the academic quality of the chosen school into a component due to different preferences and a component due to differences in school quality across FCS. We show that the constraints account for two thirds of the overall observed difference. These constraints are largely driven by admissions criteria (principally the proximity criterion), which means that choice is restricted for some households.

These results add to the live academic debate around school choice. Although there has been a great deal of work on the impact of school choice, investigating preferences directly is less common. Closest to this study, Hastings *et al.* (2008) estimate a mixed-logit demand model for schools using school choice data from Charlotte-Mecklenburg, North Carolina. They find that parents value proximity and academic attainment highly and that the preference attached to a school's mean test score increases with student income and own academic ability. They also show considerable heterogeneity in preferences. Hastings *et al.* (2008) use their model to estimate the elasticity of demand for each school with respect to mean test scores in the school. They find that demand at high-performing schools is more responsive to increases in mean test scores than demand at low-performing schools. Schneider and Buckley (2002) use an online schools database in Washington, DC to monitor parent search behaviour for schools as an indicator of parent preferences. They find that patterns of search behaviour depend on parent characteristics and find a strong interest in learning the demographic characteristics of a school's students. Rothstein (2006) adopts a more indirect approach to evaluate the relative weight parents place on school effectiveness and peer group. He finds little evidence that parents focus strongly on school effectiveness. Other authors have focused on the availability or intelligibility of the information given to parents, which may influence the choices of parents from different social groups (Ball *et al.*, 1996; Hastings and Weinstein, 2008). A number of educational and sociological studies have explored the process of parental choice in detail, often focusing on parents' stated preferences for schools (see our companion study, Burgess *et al.*, 2011).

Hastings and Weinstein (2008) make an important distinction between a family's preferences for school characteristics and the information they are able to access about the schools. Using a mix of field and natural experiments, they show that the provision of additional information on school characteristics does change school choices, particularly for disadvantaged families. In England that distinction is still

valid, but empirically it is much less of an issue. Comparative school performance data have been published since 1996 for primary schools and this is prominently reported by national and local newspapers, TV and radio, and is also supplemented by information distributed by LAs. Moreover, head teachers are very aware of parental scrutiny of the information: in a recent study all but one head teacher stated that parents are a key stakeholder group who take an interest in the school league tables (Wilson *et al.*, 2006). While there is likely to be a social gradient in the ability to process and comprehend this information, in the English context, differences in information are of second order relative to those reported by Hastings and Weinstein (2008).

In the next Section, we present our modelling framework and describe the systems by which children are allocated to schools in England. We also set out the econometric model. Section 2 details the data, focusing in particular on the definition and estimation of the FCS of schools. Section 3 presents the results, first describing the properties of these choice sets, then the conditional logit model, its sensitivity to choice set used, the results from the fixed-effects approach and the interpretation of the results. In Section 4 we offer some conclusions for policy on school choice and educational inequality.

1. Model

Below we set out our economic model of parents' choice of primary school, how we address identification issues and our econometric specifications.

1.1. *Economic Model*

Our main variable of interest is the academic quality of the school that parents nominate as their first choice on the appropriate application form. This nomination process involves: a set of schools that the parents choose from; information on the characteristics of those schools; parents' preferences over those characteristics and an allocation mechanism that maps from the parents' nomination to actual school attended by their child. Our approach is to define precisely the set of schools that parents can choose from and then, given that and the known allocation mechanism, to make inferences about parents' preferences for school attributes from the choice of school we observe. We discuss these assumptions in turn.

We assume that parents prefer the school that maximises their utility. We follow an additive random utility framework in which we assume the classical model of the rational, utility-maximising consumer (McFadden, 1977). The parent faces a choice of schools indexed $s = 1, \ldots, n$. Each parent i derives utility U_{is} from each school s, which may in general depend on characteristics of the parent, pupil and school. This is standard, and for each individual i is written $U_{is} = V_{is} + \varepsilon_{is}$, where s is the number of schools ranging from $s = 1, \ldots, n$. V_{is} denotes the deterministic component of utility and ε_{is} denotes the random component. Deterministic components include parent and school characteristics, and may be written as follows:

$$U_{is} = \mathbf{X}_{is} + \mathbf{W}_i + \varepsilon_{is} \quad s = 1, \ldots, n, \tag{1}$$

where $\mathbf{X}_{is}$ represents the characteristics of the schools (varying by alternative school) and $\mathbf{W}_i$ represent the invariant characteristics of the parent and child. Random components ε_{is} include idiosyncratic tastes of the parent and unobserved characteristics of the school choices, such as school ethos or leadership. Probabilistic statements about the distribution of choices can be made, where parent i nominates the school that maximises her utility. The probability that school s is parent i's most preferred school is as follows:

$$P_{is} = \Pr(U_{is} > U_{it}) \quad \forall t \neq s. \tag{2}$$

This preferred school will not necessarily be the one that parents actually nominate as the first choice, as the optimal nomination decision also depends on the mechanism used to assign pupils to schools, which is truth revealing only under certain conditions, which we describe next.

1.2. *The School Admissions System in England*

In England, the term 'school choice' describes parents' right to express a preference for the school they would most like their child to attend. Parents must complete a common application form to their school district/local authority (LA), on which they may nominate at least three schools. The LA then assigns pupils to schools based on these nominations, school admission criteria and the availability of places. Parents' preferences do not guarantee a place at their desired school, and school admission criteria play an important role.

Within a local authority, there are various types of school. Although most schools follow the standard admission criteria set by the LA, others set their own admission criteria. Even these schools are bound by legislative restrictions on the types of criteria that may be applied.[2] Around three quarters of primary schools in England are schools to which admission is controlled by the LA according to a published set of rules and an assignment mechanism. Criteria for entry into schools that control their own admissions are also published. Parents apply to all these different types of schools (except private schools) through a common application form.[3] We therefore observe the choices of all parents who apply to state schools, regardless of the type of school they wish their child to attend. Priority in admission is given to pupils with certain characteristics: children with a statement of special educational need; children who are looked after by the state (LA) and children with siblings who already attend the school. After these pupils, priority is generally determined by proximity to the school. The system uses a continuous measure of straight-line distance and so is rather different from the more discrete within-district *versus* outside-district criteria often used in the US.[4] For schools that control their own admissions, similar criteria are applied with the

[2] Generally, admission to Community schools (which constitute 62% of all schools in England) and Voluntary Controlled schools (which constitute 14.5%) are controlled by the school district, whereas Voluntary-Aided (mainly faith) schools and Foundation schools control their own admissions, as of course do private schools.

[3] Pupils wishing to apply to a school that controls its own admissions may also complete an additional form.

[4] In some rural areas, there are *de facto* catchment areas, i.e. a village. So children living in village A would have priority over other potential applicants who lived nearer but not in the catchment area.

added requirement that parents who wish to enrol their children in a faith-based school may have to provide proof of religiosity.

Approximately 95%[5] of all primary school-age pupils in England attend state-funded primary schools, so we focus only on these pupils and do not consider private school pupils (discussed further in Section 2).

1.3. *Identifying Parents' True Preferences*

Is it optimal for parents to nominate their true preference, or is it optimal to behave strategically? Preferences from data on the nominated schools can only be estimated where parents' choices are not strategic (or information on their truly preferred school is available). In the mechanism design literature for two-sided matching problems, Roth (1984) shows that parents' nominations for school choice will be affected by the admission criteria (the mechanism design) used. In the school choice context, Abdulkadiroğlu and Sönmez (2003) set out the mechanism design approach to school assignment. Abdulkadiroğlu *et al.* (2005) apply this approach to the Boston and NYC school districts, and Pathak and Sönmez (2008) and Abdulkadiroğlu *et al.* (2009) subsequently update the design. These studies determine the properties of particular assignment mechanisms and whether they elicit true preferences from the participants. Revealing true preferences is a weakly dominant strategy in two common mechanisms, student proposing deferred acceptance (SPDA, Gale and Shapley 1962, also called student optimal stable matching) and top trading cycles.

In the year that our sample made their decision (Autumn 2004 for entry in September 2005) there were two different assignment mechanisms in use in England. Local authorities could either operate a 'first preference first (FPF)' system in which priority is given to parents who named the school as their first choice on the application form, or an 'equal preference (EP)' system in which the rank assigned to the school by the parent is not taken into account in allocating places,[6] although parents will be assigned to their most preferred school that has space. The EP system is equivalent to the SPDA algorithm and so encourages families to reveal their true preferences. In both EP and SPDA mechanisms, families rank schools, but the admissions authority (LA or school) decides which pupils to admit to which schools without discriminating against parents who did not nominate a school as their first choice. Details of the EP system in England are given in Pennell *et al.* (2006).

The difference between the EP and a pure SPDA algorithm is that under EP parents can rank only a limited number of schools. Coldron *et al.* (2008) report that most LAs (64%) invited three preferences, 8% sought four or five preferences and another 28% requested six or seven. It is therefore possible that a pupil will not be admitted to any of these schools, encouraging at least one 'safe' choice and perhaps strategic nominations. Haeringer and Klijn (2009) and Calsamiglia *et al.* (2010) show that when parents can make only limited nominations, truth telling is not optimal in some circumstances. We believe our focus on schools with an almost-sure probability of admission

[5] From the authors own calculations, based on the Independent Schools Council Census 2009 http://www.isc.co.uk/publication_8_0_0_11_561.htm and the Annual Schools Census for Spring 2008.
[6] The former is also called the Boston mechanism (Pathak, 2011).

(discussed below) avoids this problem, however. Calsamiglia *et al.* (2010) show that in a game with k choices on the form, playing truthfully up to and including the 'safe' school is weakly dominant for those families with a school with certain entry in the top k. See also Abdulkadiroğlu *et al.* (2009), discussed below. As our preferred definition of each family's FCS only includes schools to which entry is almost certain, we overcome the problem of limited choices.

In contrast, the FPF system encourages expression of a 'safe' first choice. Parents whose true preference is a very over-subscribed school are unlikely to name it as their first nominated school as they risk losing entry to the schools nominated as their second or third choice as well. Through a survey of LAs' admissions criteria for 2006/7, Coldron *et al.* (2008) find that, in 2006, a minority of LAs use this FPF system and specifically 68% of LAs used an EP system.[7]

To address the issues raised in the mechanism design literature, we employ two strategies. We also provide brief evidence on the extent of strategic preferences made on application forms.

First, to address the problem of a constrained number of choices we define an FCS of schools for each household: schools at which the family is almost certain to be offered a place because their very close neighbours already attend the school. A desirable school that is unfeasible for some reason (perhaps over-subscription and a small catchment area) will therefore be excluded from the 'high-probability' FCS; this is detailed fully below. In the SPDA mechanism, even if parents can only rank a subset of schools, if all schools are 'safe' choices then they can do no better than ranking those schools in the correct order. This result is in the appendix of Abdulkadiroğlu *et al.* (2009).[8]

Second, we confirm that our main results are robust to estimation only on pupils in the two thirds of LAs which used the EP (SPDA) mechanism, as parents' nominations in these LAs should be truth revealing.

There are two subsidiary pieces of information that strengthen the case that the nominations are truly revealing preferences. First, as part of the survey, parents are specifically asked if they made their choice strategically;[9] only 10% claimed that they had. Second, parents are also asked if there was another school that they truly preferred to their top nominated school but that they did not put down on the form. Only 7% said this was the case and gave the name of the preferred school. In these cases we substitute the 'truly preferred' school for the school nominated on the LA form.

We note two final estimation issues. First, household location is a choice and may be endogenously affected by demand for high-quality schools. Suppose a family had moved to an area with good academic schools for this reason. This would give undue weight to proximity to the school in estimation, so the true preference for academic

[7] These relate to secondary schools but LAs will use the same system for both primary and secondary.

[8] The result does not necessarily imply that families will rank their top choice first but it does mean that the nominations will be ordered correctly; that is, the rank order will be preserved. This is therefore informative about the weights families place on their choices.

[9] Parents were asked to give the reasons they chose to nominate their first-choice school. We assume that parents acted strategically in their nominations if they give 'how likely it was that [their child] would get a place' or that the 'school is a feeder school' as an important reason for their choice.

quality would appear as a preference for proximity. The extent of the bias is limited by estimating preferences for each household using the attributes of the chosen school relative to all others in the household's specific defined choice set; preferences for academic quality and distance are estimated within the endogenous set of schools. Estimated preferences for households that move close to a particular school may remain biased, however, but a small proportion of households in the Millennium Cohort Study (MCS) report moving house strategically in this way.

Second, despite a rich data set on schools, there are likely to be unobserved school factors that influence parents' choice that are correlated with attributes in our model. Suppose such a factor makes a school over-subscribed. This in turn means that only families living close by will find it optimal to express a preference for that school. Again, this will tend to produce upward bias on the estimated preference for proximity relative to other school factors. If there are unobserved factors that are preferred by particular socio-economic groups, for example, a school is particularly attractive to high-SES families, it will therefore tend to be chosen by fewer poor families. This will raise the estimated weight on socio-economic composition, even though this is in fact due to the unobserved school factor. Given our rich set of observable school attributes, however, we are relatively unconcerned about this source of bias.

1.4. *Econometric Model*

We adopt two approaches. The first allows us to estimate household preferences for multiple school attributes simultaneously using a carefully constructed FCS of schools. While we believe this takes into account observed differences in access, estimated preferences may still be contaminated by unobserved differences in access constraints. The second allows us to deal with unobserved differences too, but forces us to focus on our school attribute of prime interest: academic quality.

First, given the model in (1) and (2) we follow McFadden's approach and assume that the error terms have standard Type 1 extreme value distributions to yield a conditional logit model:

$$P_{is} = \frac{e^{\mathbf{X}_{is}\boldsymbol{\beta}+\mathbf{W}_i\mathbf{X}_{is}\boldsymbol{\gamma}}}{\sum_{l=1}^{n} e^{\mathbf{X}_{il}\boldsymbol{\beta}+\mathbf{W}_i\mathbf{X}_{il}\boldsymbol{\gamma}}} \quad s = 1, \ldots, n, \tag{3}$$

where $\mathbf{X}_{is}$ represents a vector of school attributes that varies across the choices (and between families as we include home–school distance). The vector of coefficients $\boldsymbol{\beta}$ corresponding to these attributes is common across household types. $\mathbf{W}_i$ represents a vector of household characteristics that are constant across choices, which we interact with school attributes to allow different family types to have different preferences for school attributes. Family 'types' are classified by quintiles of socio-economic status, discussed in Section 2. This allows us to detect differences in preferences for school attributes between household types.

The second approach allows us to quantify the role of preferences and constraints in determining school choice: do pupils from lower socio-economic groups choose (and attend) lower performing schools (according to academic quality) primarily because of differences in preferences or differences in feasible school choices?

To do this we switch attention to a single school attribute, academic quality, in a linear model. We compare the school choices made by households that have exactly the same set of schools with a high probability of entry (defined in the next Section) so that differences in the set of available schools are eliminated. The residential clustering in the MCS means that there are sufficient numbers of such groups; 57% of the cohort members have exactly the same set of schools with a high probability of entry as another cohort member and variation in SES quintiles within their matched group. To the extent that our data work to define schools with a high probability of entry (detailed below) is accurate, this entirely removes differences in constraints between households with different SES, leaving only differences in preferences.

We compare the following models:

$$y_{ig} = \mathbf{W}_i\boldsymbol{\psi} + \zeta_{ig}, \tag{4}$$

$$y_{ig} = \mathbf{W}_i\boldsymbol{\theta} + \boldsymbol{\mu}_g + \varpi_{ig}, \tag{5}$$

where y_{ig} is the academic quality of the school chosen by household i from the common school choice set g, $\mathbf{W}_i$ are household SES quintiles and $\boldsymbol{\mu}_g$ are choice set fixed effects. Including $\boldsymbol{\mu}_g$ in model (5) accounts for all aspects of that shared FCS as all households within each group have exactly the same set of high-probability schools available. In model (5), $\boldsymbol{\theta}$ therefore represents the relationship between household SES and the academic quality of the chosen school, across households with almost-sure access to exactly the same schools. We therefore feel secure in interpreting any differences we observe as differences in preferences. In model (4), $\boldsymbol{\psi}$ represents the relationship between household SES and the academic quality of the chosen school, across households with different sets of high-probability schools (constraints in access to high-quality schools), thus conflating preferences and constraints. Comparing the coefficients from models (4) and (5) therefore allows us to separate the impact of preferences from the combined impact of preferences and constraints in access.

The conditional logit model and the linear fixed-effects model are complementary: the first allows us to estimate preferences for multiple school attributes jointly but may still include some influence of unobserved differences in school choice sets; the second provides a straightforward way to quantify the relative roles of preferences and constraints, but restricts the focus to one school attribute only.

2. Data

Our data set combines survey information, administrative data on school characteristics, admissions criteria and allocation rules; and spatial data. We now describe these data in turn, before defining our final estimation sample.

2.1. *The Millennium Cohort Study*

Our longitudinal survey data set is the MCS, funded by the Economic and Social Research Council and a consortium of government departments, and run by the Centre for Longitudinal Studies. The MCS was sampled from all live births in the UK. We focus on those born between 1 September 2000 and 31 August 2001 to ensure the

same academic cohort and restrict analysis to those born in England, as there are differences across education systems across the UK. In England, the sample was disproportionately stratified to ensure adequate representation of deprived areas, defined as the poorest 25% of wards based on the Child Poverty Index (CPI),[10] and also areas with high concentrations of Black and Asian families.[11]

The MCS provides rich information on parents' characteristics and their relationship with the child. Our analysis mainly uses the third survey of the MCS, taken in 2006 when children in the study were 5 years old and beginning primary school, although we draw on earlier waves for parental characteristics. Details of the parent and pupil characteristics we use can be found in online Appendix Table A1. For ease of interpretation and to enable a parsimonious model specification, we summarise a number of measures of income, occupation and education in a single indicator of socio-economic status; online Appendix Table A2 gives details of the principal component analysis used to construct this.

Parents were asked to give details of all schools they nominated on their common application form, and any other schools they applied to. The survey also asked the parent to name any schools that they would have liked their child to attend 'but chose not to apply to'. We take the parents' most preferred school as the choice they ranked highest on their application form. For the 7% of parents who say they would have preferred their child to attend a different school but they chose not to apply,[12] we use this preferred school as their first choice. The data set also provides extensive information on the reasons that parents gave for their choice of school; we analyse this in a companion study (Burgess *et al.*, 2011).

There are two potential problems with these data. First, it is clear from the chronology that parents are asked the survey questions retrospectively, on average about 12 months after they would have made their choices. Parents may not recall their nominations accurately. Parents who do not get their first choice may forget that they initially made a different choice, or may have rationalised their actual school as a good choice. If this were the case, we would expect a higher fraction of parents in our data stating that their actual school is their first-choice school. But in fact our data are broadly comparable to the national picture.[13] Furthermore, in our companion study we find that on average parents of different social class are equally likely to be admitted to their first-choice school. We therefore believe that recall bias is not a serious problem.

Second, there is a discrepancy between the percentage of parents saying that they gave a preference on the form and the national data: 72% of our sample report that

[10] The CPI is defined as the percentage of children under 16 in an electoral ward living in families that were, in 1998, receiving at least one of the following benefits: Income Support; Jobseekers Allowance; Family Credit; Disability Working Allowance, and is therefore a proxy for neighbourhood deprivation.

[11] 'High' is defined as above 30% of the ward. Proportions are based on the 1991 national census.

[12] The most common reasons parents gave was because the preferred school was too far away or they were doubtful their child would get in. Other reasons given by a minority are school expenses and religious grounds.

[13] Our data show that 94.2% of parents got their first-choice school. There are no national data for primary schools on the percentage of families getting first choice. We contacted a small number of LAs and all reported at least 90%. There are national data for secondary schools in which 83.2% of applicants get their first-choice school but we expect this number to be below that for primary school places.

they made an application through the LA application form, whereas particular local authorities that we contacted reported that at least 95% of parents expressed a preference in this way. This difference could be because parents have forgotten what they did, did not recognise the questionnaire wording as a description of what they did, were confused by the system of application or simply did not want to answer the question in full.[14] There is a mild socio-economic gradient in answers to this question, shown fully in online Appendix Table A3: 20.6% of parents who report that they made no application are from the lowest SES quintile, compared to 18.9% of parents who report at least their first choice of school. Those who report not expressing a preference for their first-choice school are also disproportionately likely to be from households with low education, single parent households and households without access to a car, although the differences are relatively small. Parents making the school application for their oldest child are more likely to make a preference, perhaps as school applications are effectively a formality for younger siblings (although a necessary formality). In our analysis, we undertake a robustness check to address this issue, restricting our sample to those without older siblings only.

We impute the first choice of school for households where choice was not reported in the survey with the school the child currently attends. This is reasonable as the proportion of households that are allocated to their first-choice school is high across England. This non-random selection is likely to bias our results to some extent, although Table 6 (column 4) and online Appendix Table A6 show that our conclusions are robust to excluding these households.

2.2. *Schools Data*

Detailed information on schools is essential to compare characteristics of the chosen school with others that were feasible. We use two administrative data sets: EduBase and the NPD.

EduBase has information for all educational establishments (state and private) in England and Wales, including the type of school, phase of education (primary, middle or secondary) and exact location (postcode). Our sample includes all non-special schools in England. We drop schools which do not admit pupils at 4/5 years old, as these schools would have been unavailable to our sample at the time we observe them.

The NPD is an administrative data set which contains information on pupils' attainment throughout their school careers, as well as information from the pupil-level census such as each pupil's eligibility for free school meals (FSM) (an indicator of poverty), recognition of any special educational need and whether they are recorded as having English as an additional language. Providing this information for the census (previously known as PLASC, now the Annual Schools Census) is a statutory requirement for all state-funded schools in England; data should therefore be accurate and reliable. We collapse the pupil-level data to school level, yielding the percentage of pupils in the school with each characteristic.

[14] The questionnaire routing and wording is available from the MCS survey website: http://www.cls.ioe.ac.uk/shared/get-file.ashx?id=750&itemtype=document

From the attainment data, we construct a measure of each school's average examination points score (averaged over 3 years; 2003, 2004 and 2005).[15] We adopt this measure of performance as it is easily observable to parents and readily understood. As discussed above, it is a very safe assumption that all parents would be aware of this information. While we cannot replicate the analysis of Hastings and Weinstein (2008) distinguishing the availability of information from preferences, as we have already indicated, we are confident that most parents in England who are about to choose a school have access to this information.

Our test score measure is an indicator of raw school attainment and so it incorporates the effect of school composition as well as teaching quality. An alternative would be for us to compute an estimate of value added as the school quality attribute. This would better represent school effectiveness but is potentially more difficult for parents to understand and it is less obvious that all parents would use this information in making their choices (Wilson *et al.*, 2006). Hastings *et al.* (2008) also show that a value-added measure does not fit their data as well.

We impute academic performance for some schools as they are split between an infant school (age 4–7) and a junior school (8–11) and we need to impute the age 11 test score outcome for the age 4 entry school.[16]

2.3. *Spatial Data: Generating Feasible School Choice Sets*

The final key element in our data is to link the families to the schools that they could have chosen; that is, to define their FCS of schools. The choice set that a household considers cannot be known, so we have to recreate it. There are no legal restrictions on the schools that a family can apply to, so in principle the feasible set of schools could be very large indeed. In the year our cohort applied for school places, parents could apply to any school inside or outside their LA, although making an application to a school outside their own authority was slightly more difficult.[17] We take two approaches to define a FCS.

Our preferred definition, the high-probability FCS, incorporates the likelihood of being offered a place at a school. In England, proximity to the school is usually the key tie-breaker in determining admission to over-subscribed schools and is widely included

[15] These are the key stage 2 (KS2) examinations, compulsory for all pupils in state schools in English, mathematics and science. These examinations are nationally set and remotely marked. They are (because of their publication) high stakes for the schools but not for the students. They do not affect school assignment.

[16] These split schools are typically closely linked (e.g. on the same site) but this is not noted in the administrative data and so has to be recreated statistically. This is necessary to attach the appropriate age 11 attainment data to the age 4 or 5 entry choice. We impute KS2 scores for each infant school based on the schools' characteristics. We also impute KS2 scores for schools missing academic information for other reasons. We impute KS2 mean scores for 3,846 of 22,324 primary schools and the proportion with level 5 in all KS2 tests in 2003 for 5,344. Another feasible method would use scores from the junior school that most pupils in an infant school attend. We explore this method by matching infant schools with their modal junior school. We compare attainment in both methods and find a high degree of correlation.

[17] In most cases, parents must submit a separate application form to the LA. Pupils in London submitted a common application form for all LAs, however. The Pan-London co-ordinated admissions system was introduced from 2005 (the year after our cohort applied to primary school). The aim of this scheme was to co-ordinated admissions between all school districts in London and some surrounding districts to reduce the number of parents receiving multiple offers, and those receiving none, see http://www.londoncouncils.gov.uk/children/briefings/PanLondoncoordinatedadmissionssystem.htm.

in admissions procedures.[18] We use the school census data to construct a measure of each school's *de facto* catchment area, the area within which prospective pupils are very likely to be admitted to the school,[19] as follows:

Calculate the straight-line distance between the home address and school attended for all pupils in England in the reception year group in 2004, the entry cohort before the MCS cohort started school. The calculation is based on the previous year's entry as this is the year that is most relevant to prospective parents in our sample.[20]

Collate this distance by school and calculate the 80th percentile of the distribution, to exclude outliers. Thus 80% of pupils attending school *S* live no further away from the school than that distance. We assume that any family contemplating choosing school *S* and living closer than that distance would feel almost certain of being offered a place at that school. That school choice could be considered a 'high-probability' option.

The probability of entry (within some tolerance) is likely to be known by parents. Most LAs provide information on the number of applications for entry and the number of places possible to be allocated in the previous year, which gives an indication of relative supply and demand. This is presented alongside other school information in the school application booklet and so is easily accessible. Some LAs also give further information on over-subscribed schools, for example how far away the furthest pupil lived from the school in the previous year. Each family's high-probability FCS consists of all schools they nominated on the form, schools for which they live within the 80th percentile of the distance distribution, which are in the LA in which they live and are within 20 kilometres of the family home.[21]

An alternative definition of the FCS is all schools that are a reasonable (although arbitrary) distance from the home. We calculate the straight-line distance between each pupil in our sample and every school in England. We assign all schools that are within 3 kilometres of the pupil's home and in the same LA to their distance FCS; we also include all schools nominated by the pupil's family. The distance FCS captures nothing about the probability of admission at each of these schools, which might be low for some schools, and disproportionately low for the probability of admission to 'good' schools.

In the fixed-effect model results presented below, we focus on households which share exactly the same high-probability FCS with other MCS households. To be clear, this is not simply having some specific schools in common in their high-probability FCS but having exactly the same set. The high spatial clustering of the MCS means that this is a surprisingly common outcome; 65% of the cohort members share their high-probability FCS with at least one other cohort member. Online Appendix Table A8 compares this sample to the full estimation sample. There are some statistically significant differences between the samples, with the sub-sample being marginally

[18] See Annex A of West *et al.* (2009).

[19] Some LAs may have official catchment areas for some schools which are non-linear. Catchment areas in rural areas may exclude one village for example. Our approach does not account for this, but is a good approximation.

[20] In fact, the correlation between years is quite strong. The same calculation in the previous-year group has a correlation with the current-year group of 0.76.

[21] This final restriction is simply to ensure that schools with very large catchment areas, such as boarding schools, are excluded from the FCS.

more disadvantaged. This implies that our results hold across a somewhat more disadvantaged group of households.

Distance is likely to be an important variable in parents' decision making. We therefore need a measure of distance between each family and each school in its distance and high-probability FCS. Calculating the distance *per se* is part of the procedure outlined above, but confidentiality requirements mean that we cannot retain actual distances for the analysis. We therefore transform the distance variable to the distance rank of schools and use that.

Finally, we map the information regarding each school district's assignment mechanism (either EP or FPF) on to the survey data.[22]

2.4. *Our Estimation Sample*

Our final sample is derived from the MCS for England, matched to the National Pupil Database (NPD) described below, yielding 9,369 households. The biggest source of missing data is families not reporting their first choice of school, reducing the sample size by around 2,535 families. We impute choice for these households with the current school attended (as described above) to maintain the sample size, but our results are robust to the exclusion of this group. We drop 12 pupils who attend special schools, as it is likely that parents' preferences might be non-standard.[23] We drop households where the child attends a private school, and we do not include such schools in the choice set for parents. We find that only 89 parents (under 1% of the sample) applied to the local authority for state education if they currently attend a fee-paying school, suggesting that the decision to apply for a state school place is taken after the decision to attend state or private education: we therefore model parents' preferences for state school attributes, conditional on the decision to apply for a state school.[24]

In summary, combining all these data sets, we have 7,905 households for the analysis of preferences and 6,048 households for the analysis of shared high-probability choice sets.

3. Results

We first characterise the distance and high-probability FCS described above. Second, we give the results from the conditional logit model: parents' preferences for multiple attributes of schools, and a test for heterogeneity in preferences across SES groups. Third, we present the results from the choice set fixed-effects model, which allows us to differentiate robustly between the role of preferences and the role of constraints in determining school choice. We finally assess the robustness of the results in

[22] Having done this, we must anonymise the school district/local authority variable for confidentiality reasons so we cannot identify pupils' region of the country.

[23] Note that we do not drop all pupils who have some level of emotional or behavioural difficulty.

[24] Our survey results show that 4.54% of pupils in our sample attend a private school, compared to around 5% nationally.

subsection 3.4. In all cases, standard errors are clustered by survey stratum to account for local clustering.

3.1. *Characterising the Feasible Choice Sets*

Table 1 shows the considerable variation in the number of schools available in families' distance and high-probability FCS. The number of schools available in the distance choice set depends on population density as there are more schools in urban areas. In the high-probability choice set, it also depends on the relative popularity of schools in the pupil's area, as the *de facto* catchment area of popular schools will shrink. As expected, Table 1 shows that the consideration of proximity-based admissions greatly restricts the number of schools available to parents.

Focusing on schools that are feasible in the probability of admission, Table 2 shows that there are also clear differences in the quality of schools available to different types of parent. High-SES families have greater access to higher quality schools and to schools with fewer FSM-eligible pupils. For example, the overall average quality score is 27.6, bottom SES quintile families have an average of 27.3 and top quintile families an average of 28.2, a difference of just over half a standard deviation. There are similar

Table 1

Number of Schools in the Distance and High-probability Feasible Choice Sets (FCS)

Number of schools	Distance FCS		High-probability FCS	
	%	Cumulative %	%	Cumulative %
0	2.32	2.32	2.71	2.71
1	4.67	6.99	4.74	7.45
2	3.29	10.28	11.41	18.86
3	3.29	13.57	15.25	34.11
4	4.18	17.75	15.90	50.01
5	3.54	21.29	13.82	63.83
6	4.71	26.00	11.00	74.83
7	4.05	30.05	8.81	83.64
8	3.75	33.80	5.68	89.32
9	3.82	37.62	3.93	93.25
10	3.89	41.51	2.26	95.51
11	4.10	45.61	1.80	97.31
12	3.80	49.41	1.73	99.04
13	4.09	53.50	0.57	99.61
14	3.37	56.87	0.26	99.87
15	3.08	59.95	0.05	99.92
16	3.99	63.94	0.03	99.95
17	3.46	67.40	0.03	99.98
18	2.85	70.25	0.00	99.98
19	2.33	72.58	0.00	99.98
20+	27.42	100.00	0.01	99.99

Notes. The Table shows percentages. The number of schools in the high-probability FCS is the number of schools for which the pupil lives within the schools' catchment area and so has a high probability of admission. Catchment areas are defined by the straight-line distance in which 80% of pupils in the previous cohort at the school lived. The number of schools in the distance FCS is the number of schools within a 3 kilometres straight-line radius of the pupil's home. The sample includes those that did not report their first choice of primary school during the survey, which has been imputed using their current school.

Table 2

Mean Characteristics of All Schools in the High-probability Feasible Choice Set (FCS), Split by Different Parent and Area Characteristics

School attribute	All	Low SES	High SES	English as additional language	English as first language	Metropolitan	Sparse town/village
Academic quality: KS2 (points) score over 3 years	27.64	27.29	28.19	27.24	27.77	27.5	28.29
	(1.63)	(1.68)	(1.46)	(1.72)	(1.58)	(1.71)	(1.30)
Free school meals (fraction of pupils)	0.20	0.27	0.13	0.32	0.17	0.28	0.07
	(0.17)	(0.18)	(0.14)	(0.18)	(0.15)	(0.18)	(0.07)
English as additional language (fraction of pupils)	0.18	0.21	0.12	0.42	0.10	0.30	0.01
	(0.26)	(0.27)	(0.20)	(0.31)	(0.18)	(0.29)	(0.02)
White (fraction of pupils)	0.78	0.73	0.85	0.49	0.87	0.62	0.98
	(0.30)	(0.31)	(0.23)	(0.33)	(0.22)	(0.33)	(0.05)
Any special educational need (fraction of pupils)	0.20	0.21	0.18	0.21	0.19	0.20	0.17
	(0.10)	(0.10)	(0.09)	(0.10)	(0.10)	(0.10)	(0.08)
Faith school (binary indicator)	0.49	0.49	0.50	0.51	0.48	0.48	0.58
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.49)
Distance rank (1 is the closest school)	6.74	7.72	5.82	9.63	5.82	8.31	3.16
	(7.64)	(9.15)	(5.91)	(10.72)	(6.06)	(9.31)	(2.34)
Catchment area (metre)	2,450.43	2,123.46	2,909.35	1,978.58	2,601.5	1,995.21	4,688.64
	(2,129.08)	(1,787.37)	(2,435.50)	(1,588.64)	(2,254.39)	(1,455.87)	(3,123.56)
N	8,858	1,633	1,699	1,565	7,293	3,244	1,594

Notes. Standard deviations are shown in parentheses. The high-probability FCS is the number of schools for which the pupil lives within the schools' catchment area and so has a high probability of admission. Catchment areas are defined by the straight-line distance in which 80% of pupils in the previous cohort at the school lived. The sample includes those who did not report their first choice of primary school during the survey, which has been imputed using their current school.

(sometimes bigger) variations in the characteristics of available schools by whether the household speaks English as an additional language. Schools in metropolitan and sparse towns/villages areas have different proportions of pupils eligible for FSM, identified as having English as an additional language, who are white and large differences in academic performance, on average.

Ignoring the constraint of the availability of schools in the distance FCS overstates the degree of choice available to parents, and biases the approximation of the available choices.

3.2. *Household Preferences for School Attributes*

We first discuss the results from the conditional logit model (3) which allows us to model parents' preferences for multiple school attributes jointly. We focus our discussion on the FCS that incorporates the likely accessibility of schools for each household: the high-probability choice set. Table 3 shows the results of a single regression, allowing the preference for each school attribute to vary by parents' SES group (with interaction terms for each group shown in different columns). The derived elasticities from each model and their bootstrapped standard errors are shown in Table 4 and discussed below.

Table 3 shows that parents' choice of school is significantly influenced by many school attributes: the school's academic quality, the proportion of pupils eligible for FSM at the school, the proportion of pupils with English as an additional language, the proportion of pupils with any special educational need, religious denomination and distance rank, all influence the probability that a school is chosen from the high-probability FCS. This finding suggests that school choice has the potential to increase academic standards in schools in England: as parents value academic standards, schools should compete on this measure to attract pupils. It is clear that other attributes are also important to parents. However, parents from each SES group prefer schools that are closer (with slight diminishing marginal returns), which suggests that parents make a trade-off between distance and academic quality. Parents' choice of school is also affected by the proportion of pupils at the school eligible for FSM (positively for lower SES groups and negatively for higher SES groups), which would limit the potential for schools to increase the number of pupils it attracts.

There is significant variation in demand across SES groups for academic quality, the proportion of pupils eligible for FSM and distance rank, shown in Table 3 by the joint significance of the interaction terms. This implies that the majority of parents across SES groups prefer schools with a higher proportion of pupils with English as an additional language, a lower proportion with any special educational need and without religious denomination, but parents across SES groups have more distinct preferences for academic quality, the proportion of pupils eligible for FSM and distance rank.

Table 4 allows us to consider parents' demand for school attributes more consistently by presenting the calculated elasticity of demand for each SES group. Elasticities are calculated for each SES group using a fixed reference levels of demand and the school attribute (shown in the first and second columns of Table 4), to make

Table 3

Parents' Preferences For School Attributes: High-probability Feasible Choice Set (FCS)

	Reference group (highest SES)	Interaction: 2nd-highest SES	Interaction: 3rd-lowest SES	Interaction: 2nd-lowest SES	Interaction: lowest SES	Joint significance test	Joint significance test: interactions
Academic quality: KS2 mean points score over 3 years (SD units)	0.270	−0.150	−0.191	−0.289	−0.363	0.029	0.018
	(0.087)	(0.096)	(0.102)	(0.098)	(0.108)		
Free school meals (SD units)	−0.539	−0.043	0.260	0.598	0.939	0	0
	(0.161)	(0.162)	(0.168)	(0.165)	(0.185)		
English as additional language (SD units)	0.171	−0.019	0.192	0.152	−0.044	0.018	0.265
	(0.109)	(0.141)	(0.134)	(0.152)	(0.152)		
White (SD units)	0.233	−0.055	−0.053	−0.123	−0.272	0.226	0.386
	(0.115)	(0.145)	(0.143)	(0.171)	(0.154)		
Any special educational need (SD units)	−0.230	0.133	0.200	0.208	0.165	0.048	0.144
	(0.075)	(0.095)	(0.088)	(0.086)	(0.087)		
Faith school (binary indicator)	−0.817	0.061	−0.073	−0.157	−0.071	0	0.562
	(0.111)	(0.120)	(0.125)	(0.133)	(0.144)		
Distance rank	−0.644	−0.068	0.044	0.166	0.261	0	0.062
	(0.091)	(0.073)	(0.076)	(0.096)	(0.108)		
Distance rank squared	0.060	0.004	−0.003	−0.012	−0.022		
	(0.011)	(0.008)	(0.008)	(0.011)	(0.012)		
N	42,047						
No. households	7,905						

Notes. Standard errors (accounting for survey design) are in parentheses. The sample includes households that apply to state schools in England and includes those that do not report making a school choice by imputing their first choice with their current school. The high-probability FCS is all schools for which the pupil lives within the schools' catchment area and so has a high probability of admission. Catchment areas are defined by the straight-line distance in which 80% of pupils in the previous cohort at the school lived. The 'joint significance test' reports the p-value of the joint significance test for the null hypothesis that all coefficients in the row (main effect and interaction terms) are zero. The 'joint significance test: interactions' reports the p-value of the joint significance test for the null hypothesis that all the interaction terms in the row are zero.

Table 4

Elasticity of Demand with Respect to School Attributes: High-probability Feasible Choice Set (FCS)

	Initial share of demand Q	Mean school characteristic X	Elasticity of demand					
			Average	Lowest SES	2nd-lowest SES	3rd-lowest SES	2nd-highest SES	Highest SES
Share of population				0.18	0.20	0.21	0.22	0.19
Academic quality: KS2 mean points score over 3 years (SD units)	0.16	17.7	1.11	−1.38	−0.28	1.17	1.78	4.0
			(0.32)	(0.631)	(0.65)	(0.715)	(0.77)	(0.8)
Free school meals (SD units)	0.16	1.2	−0.20	0.40	0.06	−0.28	−0.58	−0.54
			(0.033)	(0.057)	(0.062)	(0.066)	(0.082)	(0.096)
English as additional language (SD units)	0.16	0.9	0.17	0.10	0.23	0.27	0.11	0.13
			(0.026)	(0.052)	(0.056)	(0.062)	(0.072)	(0.07)
White (SD units)	0.16	3.9	0.44	−0.13	0.36	0.59	0.58	0.76
			(0.121)	(0.259)	(0.279)	(0.289)	(0.318)	(0.308)
Any special educational need (SD units)	0.16	1.9	−0.14	−0.10	−0.03	−0.05	−0.15	−0.37
			(0.031)	(0.062)	(0.061)	(0.066)	(0.073)	(0.084)
Distance rank	0.16	4.0	−0.46	−0.26	−0.31	−0.48	−0.67	−0.55
			(0.027)	(0.063)	(0.057)	(0.061)	(0.062)	(0.061)

Notes. Elasticity of demand is a weighted average of the elasticity of demand for the five SES group, where the weights are the share of the group in the sample which excludes those that did not report making a choice and those with a school in the high-probability choice set only. The elasticity of demand for each group is calculated using the formula $\partial Q/\partial X \times X/Q$, which for the reference group (highest SES) is $\beta \times X \times (1 - Q)$ and for other SES groups (denoted by s) is $(\beta + \beta_s) \times X \times (1 - Q)$. The calculation for distance rank includes the squared term $(\beta + \beta_s + 2\beta_2 X + 2\beta_{2s} X) \times X \times (1 - Q)$. We compute the elasticity with respect to each continuous school attribute X using the mean of X in the same sample. The mean number of schools in the high-probability FCS is 6.1. We therefore fix Q to be $1/6.1 = 0.16$. Bootstrapped standard errors (based on 200 repetitions) are reported in parentheses. The elasticity with respect to whether the school is a faith school is omitted as it is a binary variable.

consistent comparisons across households that may have different numbers of feasible schools and different types of schools available.[25]

It is clear that, on average, parents' demand for academic quality is elastic: when academic quality increases by 1%, parents' demand increases by 1.11%. Demand is most elastic with respect to academic standards, which, as discussed above, may have positive implications for the ability of school choice and competition to improve academic standards. The elasticity with respect to the proportion of pupils with English as an additional language is also positive and significant (although much smaller in magnitude) and the elasticity of demand with respect to the proportion of pupils that are white is positive but not significant.

On average, parents' elasticity of demand is negative with respect to distance rank (where a 1% increase in distance rank is associated with a 0.46% decrease in demand), the proportion of pupils eligible for FSM (where a 1% increase is associated with a 0.2% decrease) and with special educational needs (where a 1% increase is associated with a 0.14% decrease).

Table 4 suggests that there is significant heterogeneity in demand for academic standards across SES groups, which implies that schools in different areas may face different demand-side pressure, consistent with Hastings *et al.* (2008). The calculated elasticities and bootstrapped standard errors show that the highest SES group has a significantly greater response to changes in academic quality than all other SES groups and so appears distinct. The middle SES quintiles respond similarly to changes in academic quality and the lowest SES group is significantly different, with an elasticity which is negative and elastic.

There is also significant heterogeneity in demand for the proportion of pupils eligible for FSM, where the lowest SES group have a positive and significant demand elasticity and the middle-to-highest SES groups have a demand elasticity which is negative and significant.

In comparison, there is remarkably little variation in the demand elasticity with respect to distance rank, which is negative and of similar magnitude across SES groups. The demand elasticity for the lowest SES groups is significantly lower in absolute magnitude than for the highest SES group, however, which suggests that lower SES households are at least as willing to travel to schools that meet their preferences and are accessible.

An alternative definition of the FCS for each household is all schools within a feasible travel radius. Online Appendix Tables A4 and A5 present the corresponding coefficients and calculated elasticities using this alternative definition. The estimated signs of the coefficients shown in online Appendix Table A4 are generally consistent with those shown in Table 3. The calculated elasticities shown in online Appendix Table A5 show that while the pattern is similar to that shown in Table 4 for our preferred high-probability choice set, the absolute difference between estimated elasticities for academic standards and the proportion of pupils eligible for FSM is larger. This may be because some schools included in the distance FCS are likely to be too far for some households (e.g. those without access to cars in rural areas), although

[25] See notes to Table 4 for full details of the calculation.

around 15% of our sample chooses (or attends) a school that is further than 3 kilometres from their home.[26] However, the biggest disadvantage of the distance FCS is that some schools in the choice set will not be truly feasible. For example, very popular schools are likely to have smaller catchment areas, given the ability for wealthier parents to buy property close to the school, as house prices respond to demand (Black, 1999; Gibbons and Machin, 2006). Some schools are therefore infeasible in practice for some parents, despite their close proximity. This is of particular concern given our research focus on the preferences of parents from higher and lower socio-economic groups, as the choices of disadvantaged households are more likely to be constrained. This motivates our analysis using our preferred FCS which consists only of schools that should be considered high-probability choices. Using this choice set, results are more likely to reflect parents' preferences for school attributes, rather than the combination of preferences and constraints in admission.

To the extent that there are unobserved differences in the schools available to different parents (that influence the household location decisions of parents), there will be upward bias in the estimate of preferences for distance relative to school attainment. Results from the high-probability FCS may be upward biased in general, as additional schools are likely to considered 'high-probability' in practice:[27] 28% of parents make a choice of school outside their high-probability FCS, indicating that perhaps the *de facto* catchment areas we define are conservative, or that some parents are willing to take more risks.[28] To overcome these problems to some extent, we now explore results using an alternative methodology which compares the choices of parents with exactly the same set of high-probability schools. This allows us to compare the preferences of those from higher and lower SES groups more definitely, as constraints in the probability of admission are uniform within the groups.

3.3. *Fixed-effects Model Results*

In our fixed-effects model, we compare the school choices made by households from higher and lower socio-economic groups that have exactly the same high-probability FCS. These results are less likely to be biased by constraints than the conditional logit results presented above, although at the cost of only being able to consider preferences for a single school attribute.

The key issue is the degree to which we accurately identify the choice set from which households are making their choices. In this subsection, we build on the high-probability choice sets we described above and restrict the sample to those households which have exactly the same high-probability choice set as at least one other household. This allows us to include FCS fixed effects, controlling for all aspects of that

[26] The choices of these households are included in the distance choice set.

[27] Unbiased estimates of preferences require the correct definition of the choice set for each household. If too many choices are included, as is likely to be the case when using the distance FCS, then the results are likely to be downward biased, whereas if too few choices are included, the results are likely to be upward biased. This follows from (3), where the choice probabilities are affected by the number of alternatives in the denominator.

[28] Note that the proportion that make a choice outside their high-probability FCS is very similar in areas which operate 'equal preferences' and 'first preferences first' allocation mechanisms.

common set, observed and unobserved. The argument is that if the schools we identify are in common, then any that we miss from the choice set are also likely to be in common, in which case there will be no SES bias in the estimates. This is unlikely to be completely true in all cases but we believe that this is a reasonable assumption. In any case, bias will be reduced relative to results from the conditional logit model as the estimated demand for academic standards does not depend on the definition of the choice set: in this case we use the high-probability distance choice set only to identify groups of households with exactly the same set of high-probability schools.

Estimation of the fixed-effects model requires variation in socio-economic status of households within group, and sufficient variation in the attributes of schools within group; online Appendix Tables A9 and A10 show that this is the case: online Appendix Table A9 shows that the within-group variation in SES group is almost as large as the between-group variation, and online Appendix Table A10 shows that 17.6% of groups contain at least one household from the lowest and highest SES groups. Online Appendix Table A9 shows that the within-group variation in academic standards is around 70% of the between-group variation in academic standards, suggesting that households have significant choice between high-probability schools.

Table 5 shows the results from the fixed-effects model described in (4), with a focus on the relationship between parents' preference for academic attainment and their SES status. The dependent variable is the academic quality of the chosen school. Column (1) includes choice set fixed effects (high-probability FCS), which absorb the differences in choice sets, leaving differences in preferences as the sole explanatory factor between SES and school quality. In comparison, column (2) presents OLS results using the same sample, so the SES coefficients conflate both different preferences and different choice sets. For completeness, column (3) presents the OLS results using the whole sample (i.e. including those that do not share their high-probability choice set with at least one other household in the same). These results are similar to those in column (2) demonstrating that the common choice set sample is reasonably representative of the population in England.

The magnitude of the SES coefficients is dramatically reduced by the inclusion of choice set fixed effects in column (1) and many family characteristics become statistically insignificant (not shown in the Table). This suggests that constraints in feasible schools drive the majority of the observed difference in preferences for a school's academic standards between those of high and low SES. The coefficient for the bottom SES quintile falls by 55% once we take account of differences in feasible schools. For the other quintiles, the fall is even greater (67%, 71% and 77%, respectively, for the second, third and fourth most disadvantaged SES quintiles). Averaging across these, we find that 68% of the differences between SES groups in the academic quality of chosen schools is due to differences in constraints and 32% to differences in preferences.

The results from the fixed-effects model indicate that while differences in preferences for school quality across SES groups are statistically significant, they account for less than a third of overall SES differences in school quality. This shows that when comparing the choices made by parents who have identical schools available to them, the apparent relationship between parental SES and preference for school quality becomes considerably weaker than in models that are less able to take account of the constraints facing parents.

Table 5

Relationship Between Household SES and Academic Quality of Chosen School: Preferences and Constraints in Choice

	(1) Sample with exact high-probability FCS match: fixed effects	(2) Sample with exact high-probability FCS match	(3) Whole sample
SES lowest quintile	−0.455***	−1.022***	−1.008***
	(0.048)	(0.051)	(0.051)
SES 2nd-lowest quintile	−0.248***	−0.764***	−0.748***
	(0.040)	(0.041)	(0.041)
SES 3rd-lowest quintile	−0.155***	−0.531***	−0.517***
	(0.037)	(0.037)	(0.036)
SES 2nd-highest quintile	−0.067*	−0.296***	−0.288***
	(0.033)	(0.036)	(0.035)
Household speaks English as additional language	Yes	Yes	Yes
Mother's ethnic group	Yes	Yes	Yes
Mother's faith	Yes	Yes	Yes
Household has access to a car	Yes	Yes	Yes
FCS-HP group fixed effects	Yes	No	No
Observations	5,544	5,544	8,866

Notes. Standard errors (accounting for survey design) are in parentheses. *p < 0.05, **p < 0.01, ***p < 0.001. The sample with at least one exact high-probability FCS match is defined as those where the list of schools in the high-probability FCS is exactly the same as the list of schools in the high-probability FCS for at least one other household. High-probability FCS groups are the sets of households with exactly the same set of schools in their high-probability FCS. Not all households have an exact high-probability FCS match. The dependent variable is academic quality of the chosen school (measured in SD units). The effect of the independent variables are therefore interpreted as the average change in academic quality (in SD units) of the chosen school between those with the household attribute (e.g. those in the lowest SES quintile) and the relevant reference group (those in the highest SES quintile). Specification 1 accounts for high-probability FCS group-specific fixed effects, which account for regional (and very local) variation in the quality of schools available.

3.4. *Robustness Checks*

Table 6 presents robustness checks for our fixed-effects model. Column (1) repeats the main results from Table 5 for comparison. Note that numbers of observations in the subsequent columns are considerably reduced as we restrict the sample in various ways. In column (2) we include only school districts using the EP pupil assignment mechanism on the grounds that this mechanism is more likely to be truth revealing (although within a set of high-probability schools the strategic preferences in a FPF area are likely to be reduced in any case). The point estimates on the parental SES variables suggest a lower variation by SES in preferences for school quality in these districts but the pattern of coefficients on these variables remains similar; parents in the highest SES quintile value school quality to a greater extent than parents from the lowest SES quintile. It is unsurprising that results are similar to our main results as around two thirds of the sample is located in districts with the EP assignment mechanism. Column (3) shows results from a specification which only includes school districts using the FPF pupil assignment mechanism. The pattern remains similar for this sample but the differences between high and low-SES parents are somewhat greater in these districts, consistent with choices reflecting strategic considerations, although the difference in magnitude across samples is not statistically significant.

Table 6

Robustness Checks for the Relationship Between Household SES and Academic Quality of Chosen School: Preferences and Constraints in Choice

	(1) Main specification	(2) EP only	(3) FPF only	(4) Non-imputed sample	(5) Non-movers only	(6) Movers only	(7) Oldest sibling only	(8) Metropolitan area only	(9) Non-metropolitan area only	(10) EP, non-mover, oldest sibling only
SES lowest quintile	−0.455***	−0.401***	−0.582***	−0.495***	−0.466***	−0.337***	−0.418***	−0.484***	−0.430***	−0.786***
	(0.0482)	(0.0542)	(0.0989)	(0.0560)	(0.0643)	(0.0901)	(0.0884)	(0.0835)	(0.0597)	(0.212)
SES 2nd-lowest quintile	−0.248***	−0.231***	−0.283***	−0.264***	−0.260***	−0.201*	−0.213**	−0.295***	−0.210***	−0.255
	(0.0400)	(0.0462)	(0.0783)	(0.0472)	(0.0538)	(0.0812)	(0.0722)	(0.0763)	(0.0468)	(0.168)
SES 3rd-lowest quintile	−0.155***	−0.112**	−0.258***	−0.173***	−0.138**	−0.134	−0.107	−0.196**	−0.130**	−0.172
	(0.0373)	(0.0427)	(0.0749)	(0.0430)	(0.0463)	(0.0865)	(0.0604)	(0.0754)	(0.0419)	(0.108)
SES 2nd-highest quintile	−0.0666*	−0.0378	−0.127	−0.0603	−0.0689	−0.0569	−0.0311	−0.00244	−0.0874*	−0.0954
	(0.0330)	(0.0383)	(0.0655)	(0.0365)	(0.0378)	(0.0767)	(0.0508)	(0.0696)	(0.0358)	(0.0826)
Household speaks English as additional language	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mother's ethnic group	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mother's faith	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household has access to a car	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FCS-HP group fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,544	3,608	1,569	3,969	3,296	1,881	1,974	1,950	3,227	792

Notes. Standard errors (accounting for survey design) are in parentheses. *p < 0.05, **p < 0.01, ***p < 0.001. The sample with at least one exact high-probability FCS match is defined as those where the list of schools in the high-probability FCS is exactly the same as the list of schools in the high-probability FCS for at least one other household. High-probability FCS groups are the sets of households with exactly the same set of schools in their high-probability FCS. Not all households have an exact high-probability FCS match. The dependent variable is academic quality of the chosen school (measured in SD units). The effect of the independent variables are therefore interpreted as the average change in academic quality (in SD units) of the chosen school between those with the household attribute (e.g. those in the lowest SES quintile) and the relevant reference group (those in the highest SES quintile). All specifications account for high-probability FCS group-specific fixed effects, which account for regional (and very local) variation in the quality of schools available.

Column (4) displays the results from using only the non-imputed data and these are almost identical to our preferred results in column (1).

Residential location may not be exogenous as parents may choose to move house to access schools that are of higher quality. In column (5), we exploit the longitudinal nature of the MCS and restrict the sample to those families who did not report moving house between the birth of the cohort child and the survey around the child's fifth birthday. The estimates are very similar, suggesting that our results are not driven by residential sorting. In column (6), for comparison, we restrict the sample to those who move house, for whom location is potentially more endogenous and again there are no significant differences relative to our main results.

In column (7), we take account of the fact that most admissions authorities give preference to prospective pupils with siblings already at the school. By restricting the sample further to those children with no older sibling of primary school age, we focus only on families where the preference given has no relation to other siblings. The coefficients on parental SES are reduced somewhat compared to the preferred specification in column (1) (perhaps suggesting that convenience in children attending the same school outweighs considerations about contemporaneous test scores), but are not statistically significantly different.

In column (8), we focus on metropolitan areas only, comparing results to non metropolitan areas (column 9). The results are similar across metropolitan and non-metropolitan areas, which suggests that the potential benefits of school choice and competition (in raising academic standards) and the potential disadvantages (in increasing social segregation across schools) are not exclusive to urban areas. Note, however, that our model only includes those households with common schools in their FCS and therefore we cannot necessarily generalise this finding across all geographical areas of England.

Lastly, in column (10), we focus on a sample for which school choice is most likely to reflect true preferences: households in EP areas, who do not move from the child's birth to around their fifth birthday and where the MCS cohort member is the oldest sibling. While some coefficients become statistically insignificant (so those in the highest SES quintile make similar school choices as those in the lower SES quintiles), the same broad pattern holds, and the much smaller sample size limits the conclusions that can be drawn.

These robustness checks using alternative samples where endogeneity of location and considerations of older siblings are not as likely confirm our main results: households of lower socio-economic status are more likely to choose a school with lower academic standards than those in the highest quintile of socio-economic status. These differences across SES are at least partly driven by differences in preferences. That said, differences in access to schools with high academic performance, rather than differences in parents' preferences, are likely to drive the majority of the unequal allocation of pupils from difference socio-economic backgrounds across schools.

4. Conclusions

We have studied parental preferences for schools in England, finding that the academic performance of a school is an important determinant of parents' choice. This has positive implications for the ability of school choice and competition to

improve academic quality across schools. Parents also value the socio-economic composition of the school and strongly value proximity to the home, however, which limit the potential positive role of choice and competition.

We have taken particular care to distinguish whether parents' choices reflect preferences or constraints driven by the school allocation mechanism which prioritises proximity to the school. We have used census and spatial data to define empirical FCS for each survey member and find stark differences in the feasible choices available to families at opposite ends of the socio-economic spectrum. The more limited access to 'good' schools accounts for the majority of the unequal distribution of pupils from different socio-economic groups across schools in England, although preferences also play a role: comparing those with exactly the same set of schools that have a high probability of admission, those from lower socio-economic quintiles are more likely to choose a less academic school. These conclusions are consistent with previous literature: Hastings *et al.* (2008) find that richer and more educated families have a stronger preference for more academic schools.

The main challenge in our analysis is to allow adequately for the constraints in parental choice, to identify genuine parents' preferences for school attributes. Our definition of a high-probability FCS is intended to capture a minimal set of schools that families can truly access. Although carefully defined, it is possible that we have wrongly included infeasible schools in a way that is correlated with socio-economic status, which would bias our results, as differential constraints in access would not be eliminated. The choice set fixed-effects results take into account unobserved differences between choice sets and provide some further reassurance that any such remaining bias is minimal. Another issue is parents' understanding of the dominant school allocation assignment mechanism (known as 'equal preferences') which should elicit parents' true preferences: it may be that families did not understand the implications of the mechanism, so the choices we observe are strategic to some extent. We do observe some differences in parents' revealed preferences across areas with different assignment mechanisms, but these differences are not statistically significant.

We conclude with a brief discussion of the policy implications of our findings. We have shown that there are big differences in the attributes of accessible schools between households of high and low socio-economic status. These constraints account for around two thirds of the observed difference in the academic standard of the school chosen. These constraints are largely driven by admissions criteria (principally proximity), which means that choice is restricted for some households.

One response to this has been to create more schools: the current and previous governments have both endorsed a policy of enabling entry of new (more autonomous) state-funded schools in England.[29] Our research indicates that the impact on access to 'good' schools for those of lower socio-economic status, under the current allocation mechanism, will depend on where they are located.

The broader implications of our results for choice in education are mixed. Most parents in our data have a strong preference for schools with high academic

[29] The previous Labour administration encouraged the start up of so-called Academy schools, whereas the Coalition government is promoting so-called Free Schools. Both school types are state-funded semi-autonomous schools.

attainment. This supports the idea that competition to meet these preferences should help to raise standards. The measure of academic attainment we use is an absolute measure of test scores, as this is what parents are most likely to be familiar with, and not an estimate of school effectiveness (that takes into account how much value the school adds). How schools try to increase their test scores, either through increasing effectiveness or manipulating their intake of pupils, is therefore another important question in the chain between parental preferences and school effectiveness. We are mindful that although selection of pupils on the basis of observable characteristics is prohibited by the School Admissions Code, work by Allen and West (2009) has indicated that manipulation of pupil intake may be an issue at least in some schools.

Our results confirm parents' preference for a school near to home. We are confident that this is a true preference and not the result of proximity enabling entry, as all schools in our high-probability choice set should be considered feasible by parents. This implies the existence of *de facto* local monopolies, not through the lack of choice but through strong preferences for proximity among parents of primary school children, perhaps due to transport costs and practical considerations of travel with young children. This suggests that distance to secondary school may be less highly valued by parents, which we will explore in future work.

The challenge for education policy in England is to harness the preference for high academic standards evident in our results to raise attainment levels generally through choice and competition, perhaps by reducing the link between home postcode and the set of schools to which access is feasible in practice. Our results also sound a note of caution. Given our finding that there remain differences in preferences for academic quality between low and high-SES households even after allowing for differences in constraints, school policy needs to be aware of and minimise the potential for sorting along social lines.

University of Bristol
Institute for Fiscal Studies
University of Cambridge
University of Bristol

Submitted: 20 June 2012
Accepted: 28 January 2014

Additional Supporting Information may be found in the online version of this article:

Appendix A. Tables A1– A10.
Data S1.

References

Abdulkadiroğlu, A., Pathak, P.A. and Roth, A.E. (2009). 'Strategy-proofness vs. efficiency in matching with indifferences: redesigning the NYC high school match', *American Economic Review*, vol. 99(5), pp. 1954–78.

Abdulkadiroğlu, A., Pathak, P.A., Roth, A.E. and Sönmez, T. (2005). 'The Boston public school match', *American Economic Review*, vol. 95(2), pp. 368–71.

Abdulkadiroğlu, A. and Sönmez, T. (2003). 'School choice: a mechanism design approach', *American Economic Review*, vol. 93(3), pp. 729–47.
Allen, R. and West, A. (2009). 'Religious schools in London: school admissions, religious composition and selectivity', *Oxford Review of Education*, vol. 35(4), pp. 471–94.
Ball, S.J., Bowe, R. and Gewirtz, S. (1996). 'School choice, social class and distinction: the realization of social advantage in education', *Journal of Education Policy*, vol. 11(1), pp. 89–112.
Black, S. (1999). 'Do better schools matter? Parental valuation of elementary education', *Quarterly Journal of Economics*, vol. 114(2), pp. 578–99.
Burgess, S. and Briggs, A. (2009). 'School assignment, school choice and social mobility', *Economics of Education Review*, vol. 29(4), pp. 639–49.
Burgess, S., Greaves, E., Vignoles, A. and Wilson, D. (2011). 'Parental choice of primary school in England: what types of school do different types of family really have available to them?', *Policy Studies*, vol. 32(5), pp. 531–47.
Calsamiglia, C., Haeringer, G. and Klijn, F. (2010). 'Constrained school choice: an experimental study', *American Economic Review*, vol. 100(4), pp. 1860–74.
Coldron, J., Tanner, E., Finch, S., Shipton, L., Wolstenholme, C., Willis, B., Demack, S. and Stiell, B. (2008). 'Secondary school admissions', Research Report No. DCSF_RR020, Department for Children Schools and Families.
Gale, D. and Shapley, L.S. (1962). 'College admissions and the stability of marriage', *American Mathematical Monthly*, vol. 69(1), pp. 9–15.
Gibbons, S. and Machin, S. (2006). 'Paying for primary schools: admission constraints, school popularity or congestion', Economic Journal, vol. 116(510), pp. 77–92.
Haeringer, G. and Klijn, F. (2009). 'Constrained school choice', *Journal of Economic Theory*, vol. 144(5), pp. 1921–47.
Hastings, J., Kane, T.J. and Staiger, D. (2008). '*Heterogeneous Preferences and the Efficacy of Public School Choice*', Combines and replaces National Bureau of Economic Research Working Papers No. 12145 and 11805. Available at: http://aida.econ.yale.edu/~jh529/papers/HKS_Combined_200806.pdf (last accessed: 24 November 2010).
Hastings, J. and Weinstein, J. (2008). 'Information, school choice, and academic achievement: evidence from two experiments', *Quarterly Journal of Economics*, vol. 123(4), pp. 1373–414.
Hoxby, C. (2003). *The Economics of School Choice*, Chicago, IL: University of Chicago Press.
McFadden, D. (1977). 'Modelling the Choice of Residential Location', Discussion Paper, Cowles Foundation, Yale University.
Pathak, P.A. (2011). 'The mechanism design approach to student assignment', *Annual Review of Economics*, vol. 3, pp. 513 –36.
Pathak, P.A. and Sönmez, T. (2008). 'Levelling the playing field: sincere and sophisticated players in the Boston mechanism', *American Economic Review*, vol. 98(4), pp. 1636–52.
Pennell, H., West, A. and Hind, A. (2006). 'Secondary school admissions in London', Working Paper, Centre for Educational Research.
Roth, A.E. (1984). 'The evolution of the labor market for medical interns and residents: a case study in game theory', *Journal of Political Economy*, vol. 92(6), pp. 991–1016.
Rothstein, J.M. (2006). 'Good principals or good peers? Parental valuation of school characteristics, Tiebout Equilibrium, and the incentive effects of competition among jurisdictions', *American Economic Review*, vol. 96(4), pp. 1333–50.
Schneider, M. and Buckley, J. (2002). 'What do parents want from schools? Evidence from the internet', *Educational Evaluation and Policy Analysis*, vol. 24(2), pp. 133–44.
West, A., Barham, E. and Hind, A. (2009). 'Secondary school admissions in England: policy and practice', available at: http://risetrust.org.uk/pdfs/secondary-school-admissions-mar-2009.pdf (last accessed: 21 July 2014).
Wilson, D., Croxson, B. and Atkinson, A. (2006). 'What gets measured gets done. Head teachers' responses to the English secondary school performance management system', *Policy Studies*, vol. 27(2), pp. 153–71.