# What parents want: school preferences and school choice

Simon Burgess

*(CMPO, University of Bristol)*

Ellen Greaves

*(IFS)*

Anna Vignoles

*(Institute of Education, University of London)*

Deborah Wilson

*(CMPO, University of Bristol)*

November 2010

### Abstract

Parental demand for academic performance is central to the view that strengthening school choice drives up school performance. We combine survey, administrative, census and spatial data to study parental preferences for schools. We model the choices made in terms of the characteristics of the schools, families and the distances involved. School characteristics include measures of academic performance and school socio-economic composition. Using a conditional logit we show that all families have strong preferences for academic performance. There are also preferences for school composition and home-school distance. Our results suggest that preferences do not vary greatly between different socio-economic groups.

Keywords: school preferences, school quality, school choice

JEL codes: I20, I21

### Acknowledgements

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 dataset. Many thanks for very helpful comments 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. Finally, we thank the ESRC for funding this project (RES-000-22-2843).

### 1. Introduction

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). Since 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 two key questions. First, what characteristics are families actually looking for in a school? Is the school’s academic attainment record important, or do other factors out-weigh it? Second, do preferences differ between families in important ways, for example by socio-economic status? Answering these questions may help explain why children from poor families are much more likely to go to academically low-scoring schools (Burgess and Briggs 2009). This allocation of pupils across schools could arise through differences in preferences or in constraints. It is sometimes argued that poorer families care most or only about proximity of home to school, whilst middle class families focus strongly on school quality. If so, this might be a substantial contributing factor to the quality differential. Disentangling constraints from preferences is essential to resolve this long-standing debate.

In this paper we assemble a unique dataset that allows us to address these questions. We use survey information on parents’ school choices plus a rich set of socio-economic and neighbourhood characteristics. We have administrative data on school characteristics, admissions criteria and allocation rules. Crucially for identification, we also have the national pupil census, with embedded spatial information which allows us to model *de facto* catchment areas around schools. The strengths of our data complement the dataset used by Hastings et al (2008) in reaching their results on parental preferences. School choice in England means the right to express a preference for a particular school, and for this preference to be honoured if possible. The process involves parents nominating a school as their first choice on the appropriate school district (Local Authority (LA)) form. The components of the decision that we model are: the set of schools that the parents are choosing 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. There are two central identification challenges in this. 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]](#footnote-1). We combine the pupil census and the spatial information to define a minimal set of schools for which they have almost-sure access for each child in the survey dataset. Defining this set of schools for each family is crucial to address the second identification challenge, determining how the school nominated by the family relates to their true preferences as a function of the allocation mechanism. This issue is informed by the mechanism design literature that has analysed agents’ optimal responses to assignment mechanisms (see Abdulkadiroğlu and Sömnez, 2003, and Abdulkadiroğlu et al, 2008). Parents’ nominations will be affected by the admission criteria (or the mechanism design) used by schools. At the relevant date for our survey, about two-thirds of LAs in England used a mechanism likely to elicit truthful preferences among almost-sure schools (Coldron et al, 2006). Given these empirical feasible choice sets, the truth-revealing assignment mechanism and the widespread availability of school information in England, we implement a conditional logit model to estimate families’ revealed preferences.

Our results show that the three main factors that families care about are academic attainment, school socio-economic composition and the home-school distance. The clearest way of reporting our results is the trade-off that families are willing to make between a school characteristic and the home-school distance. We show that, on average, a family will be indifferent between taking their child to the nearest school as opposed to the next nearest school, if the further school has higher academic attainment by 2 percentage points, relative to an average of 28.02, or 1.325 SDs of school-level attainment. Once the family is considering schools further away the difference in academic attainment required is a third lower. This makes sense: for many families the journey to the nearest school is in a five year old child’s walking distance, and a decision to go to a school further away may well mean using a car instead. For the typical family the trade off between distance and socio-economic composition of the school is much weaker, suggesting weak parental preferences for socio-economic composition. We cannot use actual distances to each school due to confidentiality reasons, but we do have the distance rank of each school in the child’s feasible choice set. With this rank information we can use the census data to compute the average distance between the nearest and next nearest school, and this is 812 metres. We can therefore reinterpret the trade-off calculation as showing that the average family is indifferent between a school 500 metres further away from home than their nearest one if it has higher attainment by 1.23 percentage points (0.82 school-level SDs).

Our econometric model allows for potentially different trade-off values for each respondent, enabling us to assess whether there is variation in preferences across different family types. We show that preferences are heterogeneous, but that the mean differences by family socio-economic status (SES) are not large relative to the variation in preferences within SES groups. All families are estimated to have a strong preference for schools with high academic standards. Our data suggest that it is not the case that poorer families do not care about the academic attainment of schools. This suggests that inequality in access to good schools is primarily due to constraints; indeed in a companion paper we report the higher number of ‘good’ schools effectively excluded from poorer families’ choice sets (Burgess et al (2009)). The difference that we find is in line with Hastings et al (2008); richer families require a smaller increase in standards to consider a school further away from their home. We also show that on average, families prefer schools with fewer poor children. For some poor families, however, this sign reverses and they exhibit a positive preference for such schools.

These results add to the live academic debate around school choice. Whilst there has been a great deal of work on the impact of school choice, investigating preferences directly is less common. Closest to this paper, 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 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 focussed on the availability or intelligibility of the information given to parents, which may influence the choices of parents from different social groups (Hastings and Weinstein 2008; Ball et al 1996). 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 paper, Burgess et al (2009)).

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. ‘League tables’ comparing school performance have been published since 1996 for primary schools and this information is pervasive and prominently reported by national and local newspapers, TV and radio, also supplemented by information distributed by LAs. For anyone choosing a school the information is impossible to miss. While there is likely to be a social gradient in the ability to process and comprehend this information, 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 3 details the data, focussing in particular on the definition and estimation of the feasible choice sets. Section 4 presents the results, first describing the properties of these choice sets, then the central preference model, robustness checks and the interpretation of the results. In Section 5 we offer some conclusions for school choice and for educational inequality.

# 2. Model

We set out our economic approach, the identification issues and our econometric model.

*a. Economic model*

Our variable of interest is 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 from the choice of school we observe. We will 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,…, n*. Each parent *i* derives utility  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 , where *s* is the number of schools ranging from .  denotes the deterministic component of utility and  denotes the random component. Deterministic components include parent and school characteristics, and may be written as follows:

 (1)

Where represents the characteristics of the schools (varying by alternative school) and  represent the invariant characteristics of the parent and child. Random components  include idiosyncratic tastes of the parent and unobserved characteristics of the 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:

 (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 assignment mechanism, which we describe next.

*b. 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 may not translate into 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 admissions criteria. Even these schools are bound by legislative restrictions on the types of criteria that may be applied[[2]](#footnote-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. Parents apply to all these different types of schools through a common application form[[3]](#footnote-3). We therefore observe the choices of all parents, regardless of the type of school they wish their child to attend. First 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 (local authority); and children with siblings who already attend the school. After these pupils, those living closest to the school are generally given priority. The system uses a continuous measure of straight line distance[[4]](#footnote-4) and so is rather different to the more discrete within-district *vs* outside-district criteria often used in the US. For schools that control their own admissions, similar criteria are applied with the 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]](#footnote-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; we discuss the empirical aspects of this further in the Data section.

*c. Identification*

There are a number of identification issues to deal with. The first is defining the set of schools that a family can actually choose from; this is a major data exercise and is dealt with in detail in the next section.

The second is a strategic behaviour issue: is it optimal for parents to nominate their true preference, or is it optimal to behave strategically? Only in the first case can we identify preferences from data on the nominated schools. The mechanism design literature for two-sided matching problems (Roth, 1991) 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 Sömnez (2003) set out the mechanism design approach to school assignment, Abdulkadiroğlu et al (2005a, b) apply this approach to the Boston and NYC school districts, and Pathak and Sömnez (2008) and Abdulkadiroğlu et al (2008) subsequently update the design. These papers 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’ system in which priority is given to parents who named the school as their first choice on the application form, or an ‘equal preference’ system in which the rank assigned to the school by the parent is not taken into account in allocating places. The equal preference (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 without reference to the parents’ rank of the school. Details of the EP system in England are given in Pennell, West and Hind (2006).

The difference between the EP and a pure student proposing deferred acceptance algorithm is that under EP parents can rank only a limited number of schools. Coldron et al (2006) report that most 64% of LAs invited 3 preferences, 8% sought 4 or 5 preferences, and another 28% requested 6 or 7. 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 (2008) and Calsamiglia et al (2009) show that when parents can make only limited nominations, truth-telling is not optimal in some circumstances and for some players. We believe our definition of a feasible set of schools avoids this problem, however. Calsamiglia et al (2009) 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 Abdulkadiroglu et al (2008), discussed below. As our preferred definition of each family’s feasible choice set only includes schools to which entry is almost certain we overcome the problem of limited choices.

In contrast the first preference first system (FPF) encourages expression of ‘safe’ choices. 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 2006/2007, Coldron et al (2006) find that, in 2006, 68% of LAs used an EP system, the remainder using FPF[[6]](#footnote-6).

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

First, we define a feasible choice set of schools for all families: schools at which the family is almost certain to be offered a place because their very-close-neighbours already attend the school[[7]](#footnote-7). A desirable school that is unfeasible for some reason (perhaps oversubscription and a small catchment area) will therefore be excluded from the feasible choice set; this is detailed fully below. This addresses the problem of a constrained number of choices. In the student proposing deferred acceptance 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 Abdulkadiroglu et al (2008)[[8]](#footnote-8).

Second, we focus on the two thirds of LAs which used the EP (student proposing deferred acceptance) mechanism, as parents’ nominations in these LAs should be truth-revealing.

Third, 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]](#footnote-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 nominated school.

Finally, we offer evidence to corroborate our claims by comparing first and second choices made under the two allocation mechanisms. For the sub-sample who report two or more choices we compare different dimensions of the quality of the schools by family characteristics. We might typically expect a “safer” first choice under the FPF mechanism than under the EP mechanism. Figure 1 provides suggestive evidence to support this. In both LA types, the richest families, in the 4th and 5th socio-economic status quintiles, nominate a higher scoring first school than second. But for the lower socio-economic quintiles (which are typically the ones who may need to consider strategic behaviour due to their geographical proximity to lower scoring schools), the pattern reverses between EP and FPF, with the first nomination being lower scoring on average than the second in FPF LAs.

We note two final estimation issues. Household location is a choice and may be endogenous for schooling. We do three things in response to this. First, our dataset records house moves, and we run our analysis on a subset of families who did not move house after the child was born. Second, we include very detailed neighbourhood controls to characterise the location. Third, we consider the likely bias from endogenous location. Suppose a family had moved to be near enough to a popular school to ensure entry. This would give undue weight to their proximity to the school in the estimation, so the true preference for academic quality would appear as a preference for proximity. Finally, despite a rich dataset on schools, there are likely to be unobserved school factors. Suppose such a factor makes a school popular and so over-subscribed. This in turn means that only families living close by will find it worthwhile nominating that school. Again, this will tend to bias up 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 in the past and currently. This will raise the estimated weight on socio-economic composition, even though this is in fact proxying for the unobserved school factor.

*d. Econometric model*

Given the model in (1) and (2) we follow McFadden’s approach and assume that the error terms are iid type 1 extreme value to yield a conditional logit model:

  (3)

where represents a vector of school characteristics that will vary across the choices (and between families). Variables for the characteristics of the choices have a common coefficient.  represents a vector of characteristics of the families that are constant across choices, such as the respondent’s highest level of education or the ability level of the child. The value of the variable is constant across school groups, but may have a different impact on the probability of choosing each school group .

In the traditional applications of the McFadden model the options that respondents face are generic; in transport studies the choice may be between travelling by “car”, “train”, or “bus” for example; in the occupational literature it may be choice of profession (Schmidt et al, 1975). In our context choices are not generic, as respondents are geographically spread and do not choose between the same specific schools. Even within the same town two respondents may have different feasible schools to choose from. To implement the model, we therefore group the schools together into generic categories based on their attainment scores, composition and faith status. These school groups are common across all families. This approach has been used in the residential choice literature, where choices are grouped into housing “bundles” (Quigley, 1976) and in health, where demand for health-care provision focuses on choice of provider type, rather than a specific choice (Thomson et al, 2008). The grouping we adopt has further practical advantages as it reduces the dimensionality; some respondents in dense urban areas have up to 100 feasible schools within a 3km radius. Finally, grouping similar schools also makes the choices more distinct (McFadden (1973)). Details of this grouping process are provided in the Data section.

# 3. Data

We have assembled a dataset combining survey information, administrative data on school characteristics, admissions criteria and allocation rules; and spatial data.

*a. The Millennium Cohort Study*

Our longitudinal survey dataset is the Millennium Cohort Study (MCS). This was sampled from all live births in the UK, over the period from 1 September 2000 to 31 August 2001. The children in the study in England form an academic cohort. We focus only on England as there are differences in the education systems in the other countries of the UK. For England, the sample was selected from a random sample of electoral wards, disproportionately stratified to ensure adequate representation of deprived areas, defined as the poorest 25% of wards based on the Child Poverty Index (CPI)[[10]](#footnote-10), and also areas with high concentrations of Black and Asian families[[11]](#footnote-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 Appendix 1. We summarise a number of measures of income, occupation and so on in a single socio-economic status measure; Appendix 2 gives details of the principal component analysis used to construct this. We also have measures of the child’s emotional development, and ability (though obviously potentially measured with error at age 4); the family’s demographics and parenting styles. We take the “parental faith” indicator from wave 1. This ensures we have a measure of parents’ faith prior to any strategic decisions they make about their choice of primary school (which may include decisions about whether they want their child to attend a faith school)[[12]](#footnote-12).

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 asks 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[[13]](#footnote-13), we use this preferred school as their first choice. The dataset also provides extensive information on the reasons that parents gave for their choice of school; we analyse this in a companion paper (Burgess et al (2009)).

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. There is therefore a possibility that parents do 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[[14]](#footnote-14). Furthermore, in our companion paper 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.

Secondly, 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 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 (the questionnaire routing and wording is available from the authors). There is a mild socio-economic gradient in answers to this question: 31.90% of parents in the lowest SES quintile say that they made no application, compared to 26.43% of those in the highest SES quintile.

We match data from the MCS for England to the National Pupil Database (NPD) described below, yielding around 9300 families. The biggest source of missing data is families not reporting their first choice of school, reducing the sample size by some 2500 families. We drop 12 pupils who attend special schools, as it is likely that parents’ preferences might be non-standard[[15]](#footnote-15). We retain pupils with at least two feasible schools. This yields a sample of 6,756; with further missing data for the regressions, we have 6,118 pupils in the estimation.

*b. Schools data*

Detailed information on schools is essential to compare characteristics of the chosen school with others that were feasible. We use two administrative datasets: 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 (zipcode)). We keep all non-special schools in England. We drop schools which do not admit pupils at 4/ 5 years old, as they are unavailable to our sample.

The NPD is an administrative dataset 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 (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.

From the attainment data, we construct a measure of each school’s average exam[[16]](#footnote-16) points score (averaged over 3 years; 2003, 2004 and 2005). We adopt this measure of performance as it is easily observable to parents, since these scores are widely reported in the media, and readily understood. It is a very safe assumption that all parents would be aware of this information. So while we cannot replicate Hastings et al’s (2010) analysis distinguishing the availability of information from preferences, we are confident that most parents in England 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 to 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. Hastings et al (2008) also show that a value-added measure does not fit their data as well.

There are two issues we must account for in the schools data. First, private schools raise a number of problems. Choosing a fee-paying school is another branch of decision-making. Most of the school-level information for private schools is not available. Private schools are not covered by the schools census, and many do not sit the exams on which our attainment data are based. Our survey results show that 4.54% of pupils in our sample attend a private school, compared to around 5% nationally[[17]](#footnote-17). 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. We therefore drop families where the child attends a private school from our sample, and we do not include such schools in the choice set for parents. We report more on this issue in Appendix 3. Second, we have to impute academic information 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[[18]](#footnote-18).

*c. 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 feasible choice set (FCS) of schools. These define minimal sets: our aim is that all schools included in the set are truly feasible for the family, whereas it is likely to be the case that schools outside the set might also have been considered. The choice set that a family considers cannot be known, so we have to attempt 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, though making an application to a school outside their own authority was slightly more difficult[[19]](#footnote-19). We take two approaches to define a FCS.

FCS-T represents the choice set of schools that is feasible in terms of distance to travel. We calculate the straight line distance between each pupil in our sample and every school in England. We assign all schools that are within 3km of the pupil’s home and in the same LA to their FCS-T; we also include all schools nominated by the pupil’s family. The FCS-T captures nothing about whether the family might be sure of getting a place at these schools.

Proximity to the school is usually the key tie-breaker in determining admission to over-subscribed schools; it is very widely included in admissions procedures.[[20]](#footnote-20) We define a narrower, constrained choice set to capture this likelihood of not being offered a place, FCS-C. To implement this, we use the 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[[21]](#footnote-21). The process is as follows:

1. We 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[[22]](#footnote-22). This tells us how far pupils travel to attend each school.
2. We then collate that by school and calculate the 80th percentile of the distribution of distances for each school. We take 80% 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 no further away than that distance, would feel almost certain of being offered a place at that school on proximity grounds. That school choice could be considered a “safe” option.

This sort of information is very 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 oversubscribed schools, for example how far away the furthest pupil lived from the school in the previous year. Under this second approach, each family’s FCS-C 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 20km of the family home[[23]](#footnote-23).

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 FCS-T and FCS-C. 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: whether each school in the feasible choice set is the closest, second closest, third closest and so on. The representation of distance in the estimation is not straightforward and we discuss alternatives in Appendix 4.

Finally we map the information regarding each school district’s assignment mechanism (either EP or FPF) on to the survey data[[24]](#footnote-24).

*d. Defining groups of schools*

We implement the school groups defined above using the following criteria: the within group variance in school characteristics should be smaller than the between group variance; schools in the same groups should be seen by the parent as similar, while schools in different groups should have quite different characteristics. The process also helps to ensure that school groups are seen as distinct by the decision maker (McFadden, 1973).

All groups are defined relative to each family’s feasible choice set. This means that although actual schools are different for each family, each family will have a group of “relatively highly performing” schools and a group of not so highly performing schools. We define a “high-scoring” school as a school with average age 11 tests score (KS2) performance over 3 years that is greater than the median level within the FCS. A “low scoring” school is a school with an average score below the median. When schools are equal to the median value we randomly assign them either above or below. We define a school as “relatively poor” where the proportion of pupils eligible for free school meals (FSM) is above the median level within the FCS and “relatively rich” where the proportion is below the median. Again, in cases where the school equals the median it is randomly assigned. We also distinguish between faith and non-faith schools. We group schools based on relative levels of age 11 test scores (KS2) and FSM within each FCS and on their faith status. This provides groups such as “rich, low scoring non-faith” schools, “rich, high scoring non-faith” schools, and “rich, high scoring faith” schools. Appendix Tables 5 and 6 give more details of these school groups. Three quarters of families have a poor, low scoring non-faith school (group 3) in their FCS, and over half have a rich, high scoring non-faith school (group 2). The most common combinations of schools available are groups 2 and 3, and groups 2, 3 and 6 (rich, high scoring faith school).

# 4. Results

We first characterise the feasible choice sets, and the school groups. We then set out the main estimation results, followed by a series of robustness checks. We finally quantify the results in more interpretable ways.

*a. Characterising the feasible choice sets*

All pupils with at least two schools available in their FCS-C are included in the model. Table 1 shows the considerable variation in the number of schools available in families’ FCS-T and FCS-C. The number of schools available depends on population density as there are more schools in urban areas. It also depends on the relative popularity of some schools in your area as these may be excluded from your feasible choice set. The comparison of the number of schools available between FCS-C and FSC-T in Table 1 shows that catchment areas for schools greatly restrict the numbers of schools available to parents. By ignoring this constraint we would overestimate the ‘choice’ available to parents. When we focus on school groups in Table 2, less than one percent of pupils have all school groups in their FCS-C, while the modal number is 4.

Table 3 shows that there are clear differences in the quality of schools available to different types of parent. On average, schools available to parents have 16% of pupils with FSM. This is 22% for parents in the lowest SES quintile however, and 11% for parents in the highest SES quintile. Schools available to the population as a whole have 19% of pupils achieving a high standard in their age 11 (KS2) tests, 17% in schools available to parents in the lowest SES quintile and 23% for parents in the highest. Similar variations in the characteristics of schools available exist between those parents who have no qualifications and those with at least a degree. While schools in metropolitan and non-metropolitan areas have different proportions of pupils eligible for free school meals, identified as having English as an additional language, and who are white, academic performance is the same on average. These are important differences to account for in our model, and this is the motivation for defining school groups relative to the specific feasible choice sets of individual families. An analysis of the standard deviation of school characteristics (available from the authors) shows that on average all types of parents have a large degree of choice in their feasible choice set. Parents have a degree of choice in the characteristics of the school they apply to within their feasible set of schools, and this enables us to genuinely elicit parents’ revealed preferences.

*b. Estimation results*

We first discuss the results of the regression estimating the conditional logit model from (3) for all local authorities, presented in Table 4. All the results are displayed as odds ratios. The top panel of Table 4 displays the school group-specific variables. The middle panel presents some of the family-specific variables in rows, with school-group-specific coefficients in columns. We present only a subset of the coefficients in the table; other variables included in the model are listed in the footnote to the table and the full results are available from the authors. The final panel presents some diagnostic statistics for the model.

The top panel shows that the school characteristics that parents most care about are: distance rank, negatively valued but the cost per rank decreases at higher distances from home; academic achievement, positively valued and concave; and school socio-economic composition (the proportion of children from poor families), negatively valued and linear. We postpone discussion of the quantitative significance of these coefficients. The results show that families do not value a school by its proportion of pupils with English as an additional language, ethnic composition or the proportion of pupils with special needs.

Each column represents a school-group. In each column, the coefficients in the middle panel are interpreted as the impact of a variable on the likelihood of choosing a school in that group relative to a school in the omitted group (the “poor, low-scoring, non-faith” school-group). The pattern of the socio-economic status odds ratios suggests a preference for non-faith schools. It also suggests that richer families are more likely to choose a school in the higher-scoring school-group over the omitted group than the poorest fifth of families. The overall pattern is summarised in Figure 2: high SES families are more likely to choose schools in the “rich, high-scoring, non-faith” group and less likely to choose schools in the “poor, low-scoring, non-faith” group and, in the top two quintiles, marginally less likely to choose “poor, high-scoring, non-faith” schools[[25]](#footnote-25). The strongest effects are for the 2nd and 3rd SES quintiles being much more likely to choose “poor, high-scoring” schools over “poor, low-scoring” schools than the lowest quintile. The p-value for the joint significance of SES is 2.88%.

Parents with post-compulsory education[[26]](#footnote-26) are more likely to choose the “rich, high-scoring, non-faith” school-group compared to the reference school-group, although these are the only significant effects for parental education variables. Neighbourhood deprivation also has only a very small impact, which suggests we are successfully controlling for neighbourhood constraints parents may face in their school choice[[27]](#footnote-27). Self-reported faith has the strongest effect of the other parental variables included in the model. Recall that this is taken from the first wave of the survey, possibly before any incentive for ‘strategic’ motives for a change in faith. Unsurprisingly, self-reported faith has a significant impact on the probability of choosing a faith school-group relative to the reference school-group. There is a much greater likelihood of choosing a school in the richer faith school-groups rather than the poorer faith school groups, however, suggesting that parents value composition of the school as well as its denomination. Measures of parental interaction with the child (such as frequency of reading and visits away from home) also have significant effects.

Included factors with little or no significant effect include the ethnicity of the main parent, the availability of a car, demographics (single parent, birth order and number of siblings, gender and season of birth), and measures of the emotional and cognitive development of the child. The model includes a binary variable for whether the parent expressed a preference for a school that was further than 3km from their home. This captures a willingness to travel with the child (conditional on urban/rural location). Parents who applied to a school far from their home are more likely to apply to almost all types of school-group (except the ‘rich, low-scoring non-faith’ school-group). This suggests that parents are willing to travel for high academic standards and a religious affiliation of the school, although this variable partially reflects the larger catchment areas of faith schools more generally.

The final panel of the table gives McFadden’s *R2* of 0.144. We also calculate that the number of correct predictions of group choice from the model is 59%. This is a creditable performance and suggests that the model is capturing a good deal of the underlying behaviour. We report this statistic by SES quintile and the model provides a similar fit across SES. We report a number of tests of joint significance including socio-economic status (p-value of 7.4%), self-reported faith (0%), other parental characteristics (0%), child characteristics (27.3%) and neighbourhood deprivation (15.4%). It is reassuring to note that both the school-specific variables and the family-specific variables are significant. Unsurprisingly, given the similarity of our defined school groups, the Hausman test for the IIA property fails. One standard response to the failure of the IIA is to consider a nested logit approach. For some families this would make sense: they may first want a faith school of any type, followed by high test scores. Other families may want high test scores above all else, and so on. But because there is no natural order for the nesting, we cannot implement this and hence take the conditional logit approach.

We illustrate the importance of using FCS-C in our main regression by comparing the results with those under FCS-T (this latter choice set being determined by our 3km distance rule discussed earlier). This also serves to underline the value of our data which includes the spatial element of the *de facto* catchment areas. Table 5 presents a selection of parameters from estimating the same model but using FCS-T; the FCS-C results are repeated for easy comparison. There are a number of changes. First, the effect of school quality is significantly lower in the FCS-T model. Our interpretation of this is that families are appearing not to ‘choose’ high quality schools in the FCS-T model which are in reality unavailable, whereas in the FCS-C model those schools are correctly excluded from the choice set. There is also a lower utility cost to distance in the FCS-T model; again this is probably because families are ‘choosing’ more distant schools over local ones, whereas in the FCS-C model those nearby schools are actually not available. The results for the family control variables are much more important in the FCS-T model, particularly the SES and neighbourhood variables. Our interpretation is that neighbourhood in particular is important in describing the true choice set, and if this is conditioned on, then neighbourhood is unimportant in influencing choices. Finally, the FCS-T assumption does a worse job of predicting choices. We therefore focus on the results using FCS-C from now on.

*c. Robustness results*

In Table 6, we present the results of a number of robustness checks; column 1 re-presents the main results. Note that numbers of observations in all the subsequent columns are considerably reduced. In column 2 we include only school districts using the EP pupil assignment mechanism. The point estimates suggest a lower valuation of school quality, though not significantly so. The distance rank coefficients are very similar, suggesting it is more likely to be schools’ catchment areas rather than distance which constrain choice. The pattern of coefficients on the family-specific variables changes only marginally; the SES and parental education coefficients remain significant and become a little larger. It is unsurprising that results are similar to our main results as almost two thirds of the sample is the same. We would like to compare coefficients between samples in EP with FPF LAs but the smaller size of the latter sample coupled with a demanding estimation prohibits this.

Residential location may not be exogenous. We exploit the longitudinal nature of the MCS and restrict the sample to those families who did not report moving house since the birth of the cohort child[[28]](#footnote-28). In column 3, the school SES composition variable becomes insignificant, and the impact of academic attainment increases substantially. Socio-economic status is less important (joint significance p-value of 29.4% and very few individually significant coefficients) as are parental characteristics (p-value of 6.2%) and neighbourhood controls (24.2%). This may be partly due to the lower sample size. This suggests that when we focus on a sample where location is more likely to be exogenous, we find that school characteristics have a stronger impact on parents’ preferences but family SES differences are less prominent, implying high SES families who move house drive some of the differences in preferences.

Third, 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 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 results are in column 4. Again, the valuation of school quality is higher, and school SES composition is also valued more (negatively). Socio-economic characteristics are less important than the base case (p-value of 29.3%), parental characteristics (6.2%) and neighbourhoods (53.0%) likewise.

Overall, we argue that the main points of our estimation results are strengthened by considering these sub-samples in which endogeneity is less likely to be a problem. Whilst it would be useful to estimate the model for the sample of those in households with no older sibling of primary school age and who have not moved house in the period we observe, unfortunately this leaves too few observations for the model to converge.

We checked a number of other potential issues in the estimation, including interaction of faith school status and distance, omitting the “distant school” indicator and so on; none of these made any substantive difference to the results. As we noted above, to the extent that there are unobserved school effects, we suspect that these will tend to bias up the distance estimate relative to school attainment. Taking this section as a whole, it seems likely that our preferred specification provides a conservative estimate of the value that parents place on school quality relative to distance from home.

*d. Quantifying the effects*

We quantify parental preferences more directly by estimating the trade-offs between school characteristics. The model implies trade-offs will differ by family characteristics, directly addressing whether different families have different preferences. The non-linear form that we allow for preferences means that these trade-offs are not constants. We set this up as follows.

A family is indifferent between two schools in groups 2 and 1 if the probability of choosing each school group is the same, that is *P(Yi2 = 1)* = *P(Yi3 = 1).* Expanding this with the conditional logit model given in (3) gives[[29]](#footnote-29):

 (4)

Consider a family choosing between their nearest school (distance rank 1) with quality score *K* and the second nearest school (rank 2) with score (*K + d*); what value of *d* makes a family indifferent between the schools? To start, we simply focus on the school characteristics *X*, and reintroduce the family characteristics *W* shortly; we assume that all other characteristics of the two schools are equal. Equating the numerators of (4) for the two schools:

$Kβ\_{1}+K^{2}β\_{2}+1.β\_{3}+1.β\_{4}=\left(K+d\right)β\_{1}+\left(K+d\right)^{2}β\_{2}+2.β\_{3}+4.β\_{4}$ (5)

where *β1* is the coefficient on school attainment, *β2* on school attainment squared, *β3* on distance rank and *β4* on distance rank squared. Rearranging this yields a quadratic in *d*:

$0=d^{2}β\_{2}+\left(β\_{1}+2Kβ\_{2}\right)d+β\_{3}+3β\_{4}$ (6)

We call the solution to this the ‘core’ trade-off, which is common across all families[[30]](#footnote-30). Practically, in all the cases we considered this quadratic equation yielded a ‘sensibly’ valued root and a non-sensible root. We compute the trade-off for a comparison of schools at ranks 3 and 4 similarly. Reintroducing the family characteristics:

$0=d^{2}β\_{2}+\left(β\_{1}+2Kβ\_{2}\right)d+\left(β\_{3}+3β\_{4}-\left(con\_{1}-con\_{2}\right)-W\left(γ\_{1}-γ\_{2}\right)\right)$ (7)

where *conj* and γ*j* refer to school groups *j* = 1,2 as in (4). Equation (7) allows us to consider how trade-offs vary between parent types; the trade-off value will depend on the value chosen for *K*, family characteristics *W* and the change in the school group.

We present the results in Table 7. They show that a typical family is indifferent between the nearest school and the next nearest if the academic quality of the further school is 2 percentage points of school level academic attainment higher, relative to an average of 28.02, or 1.325 SDs of school-level attainment. Comparing the choice of schools that are further away, the difference in academic attainment required is a third lower (1.35 percentage points or 0.894 of an SD). This makes sense – for many families the nearest school is within walking distance, while a decision to go to a school further away could mean using other transport[[31]](#footnote-31). There is a similar story with respect to the trade-off of distance and socio-economic composition. In this case, distance rank 1 to 2 is traded off against 34.5 percentage points of school-level FSM, or 2.267 SDs.

We are forced to use distance rank rather than actual distance because of confidentiality issues. We can use the administrative data to get a sense of the average scale of one distance rank in terms of kilometres, however. To do this we use all pupils in the NPD that began primary school in the year before our cohort. We identify each pupil’s nearest school, second nearest school and so on. We then compute the average distance between first and second ranked school, second and third, and so on. The details of this exercise are available from the authors. It shows that on average[[32]](#footnote-32), one distance rank is about 800m.

Combining the two, we can see that a typical family is indifferent between a school with 1 percentage point higher school-level academic attainment school-level and 17 percentage points higher FSM rate (or 1 SD of academic quality against -1.711 SDs of SES composition). A school with a high intake of low income children can attract other families if it performs well academically. Given that on average such children under-perform relative to their better-off peers, the school’s effort has to more than offset this factor to be as attractive.

The second goal of this paper is to investigate whether different families have different preferences. These differences derive from the final term in (7),$ W\left(γ\_{1}-γ\_{2}\right)$, and are common to both trade-offs. The distribution of the estimated values of this term show considerable variation both within SES quintile and across. The standard deviation within quintiles is 0.641 (and an IQR of 0.810) relative to the difference between the mean value in the poorest quintile and the mean value in the richest quintile of 0.733. Thus while there are differences in preferences (as measured by the trade-offs) between SES groups, they are not large compared to within-quintile differences.

In Figures 3a and 3b we plot the distribution of the full trade-offs of distance rank 1 versus 2 for academic quality and SES composition. We see a considerable overlap of preferences. Preferences are very similar across the better off two quintiles of families, and somewhat distinct and more diffuse in the bottom three quintiles. This is more marked for the SES composition trade-off, and among the poorer three quintiles, considerable fractions have a positive trade-off and value higher fractions of FSM children. This suggests that families may have a preference for families like themselves, with richer families preferring schools with fewer poor children and poorer families preferring schools with more poor children. It is worth emphasising that all SES quintiles have very strong positive trade-offs for school academic quality, and preferences for academic quality are generally less diffuse than for composition. The differences that there are in Figure 3a suggest that those in the highest SES quintiles require less an increase in academic standards to choose a school further away from home than those in the lowest SES quintile (in line with Hastings et al, 2008).

Our final exercise is to predict the choices of a representative low SES family facing a choice set that a high SES family would typically see; and *vice versa*. The results of this are in Table 8. The patterns reflect the discussion so far: in the low SES choice set, both the representative low SES family and the fictional high SES family would decisively choose the rich, high-scoring, non-faith school. The rank ordering is the same, and that school is the most preferred by a large margin. In the high SES choice set, again both the representative high SES family and the fictional low SES family make the same choice, but this is much more of a marginal preference for the latter, reflecting their lower disutility from choosing a school with a higher fraction of poor students.

# 5. Conclusions

We have studied parental preferences for schools in England, distinguishing between constraints and preferences. We have used census and spatial data to define and construct empirical feasible choice sets for each survey member and find stark differences in the options available to families at different ends of the socio-economic spectrum.

Our results add to the literature in a number of ways. We show that the main characteristics that parents care about in a school are academic attainment, travel distance and school socio-economic composition. While these factors are not a surprise, we are able to quantify this and also be confident that we have identified parents’ preferences not the constraints they face. We estimate the trade-offs parents are willing to make between distance and attainment, and between distance and composition. We show that families do not appear to care so much about ethnic composition[[33]](#footnote-33), or proportions of children with special educational needs. Instead all families are estimated to have a strong preference for schools with high academic standards. It is not the case that poorer families do not care about the academic attainment of schools. Preferences are heterogeneous, but the mean differences between families in different SES groups are not large relative to the variation in preferences within SES groups. Consistent with Hastings et al (2008), richer families require a smaller increase in standards to consider a school further away from their home. Preferences for the socioeconomic composition of schools are weaker but on average, families prefer schools with fewer poor children. For some poor families, however, this sign reverses and they exhibit a positive preference for such schools. It may be that what families want is for their child to go to a school with other children “like” their own in terms of socio-economic status. This would obviously encourage greater social stratification in schools.

Of course, there are caveats. The main challenge is the specification of the constraints in order to identify the preferences. Our constrained feasible choice sets are intended to capture a minimal set of schools that families can truly access. We have undertaken a number of sensitivity tests, but it remains possible that we have wrongly included infeasible schools in a way that is correlated with socio-economic status. Another issue is the “Equal Preferences” assignment mechanism: it may be that families did not truly understand it, or did not respond optimally to its truth-revealing characteristic. We note however, that when we compare the results with all families to subsets of non-house-moving families and no-elder-sibling families, the apparent preference for academic attainment of schools strengthens considerably in the results, giving us more confidence in our results.

We conclude with a brief discussion of the policy implications of our findings. In terms of educational inequality, we have shown big differences in the choice sets of different families. We have shown less important differences in their preferences. The big driver of differential access to better schools is the characteristics of schools nearby to where the families live, and the use of proximity as a tie-breaking device. This relates immediately to practical issues about the operation of the current system and ensures that for some parents choice is extremely limited due to their geographical location. This limit to parental choice has long been recognised and the previous Labour administration and the current Coalition government both endorse a policy of enabling entry of new (more autonomous) state funded schools in England[[34]](#footnote-34). Our research indicates that their impact on differential access will depend on where they are located.

The broader implications of our results for choice in education are mixed. Parents, almost universally in our data, have a strong preference for schools with high academic attainment. This supports the idea that competition to meet these preferences should help to raise standards. The measure of academic attainment we use is raw test scores not an estimate of school effectiveness. How schools increase their test scores is another question in the chain between parental preferences and school effectiveness, and it is likely that schools will both aim to raise effectiveness and to manipulate their intake peer group.

The results confirm a strong 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 FCS-C are available. This implies the existence of de facto local monopolies, not through the lack of choice, but through a preference not to travel far perhaps due to transport costs. This combines with peer group preferences to raise socio-economic sorting. The challenge for education policy in England is to harness the preference for high academic standards to raise attainment levels generally, whilst minimising the potential for sorting.

**References**

**Abdulkadiroğlu, Atila and Tayfun Sönmez.** 2003. "School Choice: A Mechanism Design Approach." *American Economic Review*, vol. 93 no. 3, pp. 729-47.

**Abdulkadiroğlu, Atila; Parag A. Pathak; Alvin E. Roth and Tayfun Sönmez.** 2005. "The Boston Public School Match." *American Economic Review*. *.* vol. 95, no. 2. pp. 368 - 371

**Abdulkadiroğlu, Atila; Parag A. Pathak and Alvin E. Roth.** 2008. "Strategy-Proofness Vs. Efficiency in Matching with Indifferences: Redesigning the NYC High School Match." *Working Paper 14864*.

**Allen, R and Anne West** 2009. “Religious schools in London: school admissions, religious composition and selectivity”. *Oxford review of education*, 35 (4). pp. 471-494.

**Ball, Stephen J.; Richard Bowe and Sharon Gewirtz.** 1996. "School Choice, Social Class and Distinction: The Realization of Social Advantage in Education." *Journal of Education Policy*, 11(1), pp. 89 - 112.

**Burgess, Simon and Adam Briggs.** 2009. "School Assignment, School Choice and Social Mobility." *Economics of Education Review*. vol. 29(4), pages 639-649.

**Calsamiglia, Caterina; Guillaume Haeringer and Flip Klijn.** 2008. "Constrained School Choice: An Experimental Study." *UFAE and IAE Working Papers; 757.08*.

**Coldron, John; Emily Tanner; Steven Finch; Lucy Shipton; Claire Wolstenholme; Ben Willis; Sean Demack and Bernadette Stiell.** 2008. "Secondary School Admissions." *Research Report No DCSF\_RR020*.

**Gale, D and L. S. Shapley.** 1962. "College Admissions and the Stability of Marriage." *The American Mathematical Monthly*, 69(1), pp. 9-15.

**Haeringer, Guillaume and Flip Klijn.** 2008. "Constrained School Choice." *UFAE and IAE Working Papers; 671.06*.

**Hastings, Justine; Thomas J. Kane and Douglas Staiger.** 2008. “*Heterogeneous Preferences and the Efficacy of Public School Choice*.” Combines and replaces National Bureau of Economic Research Working Papers No. 12145 and 11805. aida.econ.yale.edu/~jh529/papers/HKS\_Combined\_200806.pdf (accessed 24.11.10).

**Hastings, Justine S., and Jeffrey Weinstein.** 2007. "Information, School Choice, and Academic Achievement: Evidence from two Experiments" Quarterly Journal of Economics, vol. 123, pp. 1373 – 1414.

**Hoxby, Caroline.** 2003. "The Economics of School Choice." *The Economics of School Choice*. University of Chicago Press, Chicago.

**Jacob, B. A. and L. Lefgren.** 2007. "What Do Parents Value in Education? An Empirical Investigation of Parents' Revealed Preferences for Teachers." *Quarterly Journal of Economics*, 122(4), pp. 1603-37.

**McFadden, Daniel.** 1977. "Modelling the Choice of Residential Location." *Cowles Foundation Discussion Papers*, Cowles Foundation, Yale University(477).

**McFadden, Daniel.** 1973. "Conditional Logit Analysis of Qualitative Choice Behaviour." *in P. Zarembka, ed, Frontiers in Econometrics*.

**Pathak, Parag A. and Tayfun Sönmez.** 2008. "Levelling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism." *American Economic Review*, 98(4), pp. 1636–52.

**Pennell, Hazel and West, Anne and Hind, Audrey** (2006) “**Secondary school admissions in London”** *Clare Market Paper, 19*. Centre for Educational Research, London, UK

**Quigley, John M.** 1976. "Housing Demand in the Short Run: An Analysis of Polytomous Choice." *NBER*.

**Roth, Alvin E.** 1984. "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory." *Journal of Political Economy*, pp. 991-1016.

**Rothstein, Jesse 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*, 96(4) pp. 1333 – 1350

**Schmidt, Peter and Robert P. Strauss.** 1975. "The Prediction of Occupation Using Multiple Logit Models." *International Economic Review*, 16(2), pp. 471-86.

**Schneider, M. & Buckley, J.** (2002) What do parents want from schools? Evidence

from the internet. *Educational Evaluation and Policy Analysis,* 24(2)**:** 133-

144.

**Thomson, Sarah and Anna Dixon.** 2008. "Choices in Health Care: The European Experience." *Journal of Health Services Research & Policy*, 11(3), pp. 167-71.

**Schools Admissions Code** (2007): <http://www.dcsf.gov.uk/sacode/downloads/DfES%20Schools%20text%20final.pdf>

Pan-London Co-ordinated admissions system:

<http://www.londoncouncils.gov.uk/children/briefings/PanLondoncoordinatedadmissionssystem.htm>

# Tables and Figures

Figure 1: Comparison of schools nominated on the LA form as 1st and 2nd



Note 1, 2, … 5 refer to quintiles of the socioeconomic status variable

**Figure 2: Predicted probability of choosing the four most common types of school, by socio-economic status of the main respondent**



Figure 3a: Distribution of trade-offs of distance rank against school academic quality



**Figure 3b: Distribution of trade-offs of distance rank against school SES composition**



|  |
| --- |
| **Table 1: Number of schools in FCS-T and FCS-C, the feasible choice set based on travel distance alone (FCS-T) and the more constrained feasible choice set (FCS-C)** |
|  | FCS-T | FCS-C |
| Number of schools | Percentage | Cumulative percentage | Percentage | Cumulative percentage |
| 2 | 4.3 | 4.3 | 8.4 | 8.4 |
| 3 | 5.3 | 9.6 | 16.6 | 25.0 |
| 4 | 6.1 | 15.7 | 16.9 | 41.9 |
| 5 | 4.7 | 20.4 | 17.0 | 58.9 |
| 6 | 6.3 | 26.7 | 13.2 | 72.1 |
| 7 | 4.9 | 31.6 | 10.8 | 82.8 |
| 8 | 4.6 | 36.3 | 7.7 | 90.5 |
| 9 | 4.6 | 40.9 | 4.1 | 94.6 |
| 10 | 59.1 | 100.0 | 5.4 | 100.0 |
| Note: The table shows weighted percentages. The number of schools in FCS-C is the number of schools for which the pupil lives within the schools' catchment area. Catchment areas are defined by the straight line distance in which 80% of pupils in the previous cohort lived. The number of schools in FCS-T is the number of schools within a 3km straight line radius of the pupil’s home. |

|  |
| --- |
| **Table 2: Number of school-groups in FCS-T and FCS-C, the feasible choice set based on travel distance alone (FCS-T) and the more constrained feasible choice set (FCS-C)** |
|  | FCS-T | FCS-C |
| Number of school-groups | Percentage | Cumulative percentage | Percentage | Cumulative percentage |
| 2 | 7.4 | 7.4 | 14.9 | 14.9 |
| 3 | 11.0 | 18.4 | 25.9 | 40.9 |
| 4 | 14.4 | 32.8 | 28.4 | 69.3 |
| 5 | 25.7 | 58.5 | 19.1 | 88.3 |
| 6 | 23.8 | 82.3 | 9.7 | 98.0 |
| 7 | 10.9 | 93.2 | 1.8 | 99.8 |
| 8 | 6.8 | 100.0 | 0.2 | 100.0 |
| Note: The table shows weighted percentages. The number of school-groups in FCS-C is the number of school-groups which are constructed from each school for which the pupil lives within the schools' catchment area. Catchment areas are defined by the straight line distance in which 80% of pupils in the previous cohort lived. The number of school-groups in FCS-T is the number of school-groups constructed from all schools within a 3km straight line radius of the pupil’s home. |

|  |  |  |
| --- | --- | --- |
| **Table 3: Mean characteristics of all schools in the narrow feasible choice set, split by different parent and area characteristics** |  |  |
|   | Mean school characteristics: by different sample splits |
| School characteristic | All | Low SES | High SES | No education | Degree + | Metropolitan | Sparse town/ village |
| Proportion of students with FSM | 0.18 | 0.24 | 0.11 | 0.26 | 0.13 | 0.25 | 0.07 |
| KS2 mean score (averaged over 3 years) | 27.70 | 27.33 | 28.25 | 27.22 | 28.14 | 27.58 | 28.31 |
| Proportion of students with EAL | 0.15 | 0.17 | 0.09 | 0.25 | 0.13 | 0.27 | 0.01 |
| Proportion of students with some SEN | 0.19 | 0.21 | 0.18 | 0.21 | 0.18 | 0.19 | 0.17 |
| Proportion of students that are white | 0.81 | 0.78 | 0.89 | 0.68 | 0.84 | 0.66 | 0.97 |
| "Catchment area" | 2403.50 | 2117.67 | 2837.03 | 1939.42 | 2864.93 | 1920.30 | 4268.64 |
| Number of schools within 3km of pupil | 18.58 | 21.32 | 14.58 | 23.57 | 16.38 | 25.28 | 4.27 |
| Note: Weighted values are shown. Statistics are calculated from all schools (as opposed to school groups) in the narrow feasible choice set (FCS-C). Details of this process can be found in appendix 2. FCS-C is defined as all schools for which the pupil lives within the schools' catchment area. Catchment areas are defined by the straight line distance in which 80% of pupils in the previous cohort lived. Those with 'low SES' are those parents who are in the lowest socio-economic quintile, 'high SES' in the highest socio-economic quintile. Socio-economic quintile was computed using principal component analysis; details can be found in appendix 4. |
|  |  |  |  |  |  |  |  |

|  |
| --- |
| **Table 4: Conditional logit: The probability of choosing a school of each school group.**  |
|  Odds ratios |   | Rich, low scoring, non-faith | Rich, high scoring, non-faith | Poor, high scoring, non-faith | Rich, low scoring, faith | Rich, high scoring, faith | Poor, low scoring, faith | Poor, high scoring, faith |
| School characteristics (characteristics of the choices): |
| Proportion FSM | 0.248\* |  |  |  |  |  |  |  |
|  | 2.10 |  |  |  |  |  |  |  |
| Proportion FSM squared | 0.974 |  |  |  |  |  |  |  |
|  | 0.03 |  |  |  |  |  |  |  |
| KS2 mean (averaged over 3 years) | 11.88\*\*\* |  |  |  |  |  |  |  |
|  | 6.40 |  |  |  |  |  |  |  |
| KS2 mean (averaged over 3 years) squared | 0.957\*\*\* |  |  |  |  |  |  |  |
|  | 6.17 |  |  |  |  |  |  |  |
| Proportion with EAL  | 1.740 |  |  |  |  |  |  |  |
|  | 1.74 |  |  |  |  |  |  |  |
| Proportion White British | 1.170 |  |  |  |  |  |  |  |
|  | 0.52 |  |  |  |  |  |  |  |
| Proportion with some level of SEN | 0.694 |  |  |  |  |  |  |  |
|  | 1.31 |  |  |  |  |  |  |  |
| Distance rank | 0.567\*\*\* |  |  |  |  |  |  |  |
|  | 21.56 |  |  |  |  |  |  |  |
| Distance rank squared | 1.035\*\*\* |  |  |  |  |  |  |  |
|  | 11.96 |  |  |  |  |  |  |  |
| Parent and pupil characteristics (characteristics of the choosers): |
| Socio-economic status: 2nd quintile |  | 1.480 | 1.526\*\* | 1.744\*\* | 0.987 | 1.198 | 0.951 | 1.070 |
|  |  | 1.91 | 2.68 | 2.96 | -0.04 | 0.99 | -0.23 | 0.20 |
| Socio-economic status: 3rd quintile |  | 1.557\* | 1.569\*\* | 1.862\*\* | 0.908 | 1.573\* | 0.994 | 1.280 |
|  |  | 2.02 | 2.69 | 2.99 | -0.29 | 2.37 | -0.02 | 0.68 |
| Socio-economic status: 4th quintile |  | 1.524 | 1.758\*\* | 1.590\* | 1.100 | 1.566\* | 0.894 | 1.518 |
|  |  | 1.80 | 3.18 | 2.07 | 0.28 | 2.23 | -0.44 | 1.14 |
| Socio-economic status: highest quintile |  | 1.952\*\* | 1.755\*\* | 1.496 | 0.817 | 1.646\* | 0.738 | 1.535 |
|  |  | 2.61 | 2.86 | 1.60 | -0.55 | 2.29 | -1.07 | 1.11 |
| Highest level of education: 'Other' or vocational |  | 1.246 | 1.757\* | 1.170 | 1.958 | 0.720 | 1.109 | 1.678 |
|  |  | 0.75 | 2.31 | 0.61 | 1.44 | -1.09 | 0.29 | 0.93 |
| Highest level of education: GCSE grades D-G |  | 1.264 | 1.101 | 0.886 | 1.740 | 0.977 | 0.937 | 1.915 |
|  |  | 0.91 | 0.46 | -0.52 | 1.30 | -0.10 | -0.22 | 1.43 |
| Highest level of education: GCSE grades A\*-C |  | 1.470 | 2.008\*\*\* | 1.097 | 1.754 | 1.333 | 1.554 | 2.297\* |
|  |  | 1.74 | 3.82 | 0.46 | 1.48 | 1.38 | 1.69 | 2.01 |
| Highest level of education: AS/A level |  | 1.310 | 2.494\*\*\* | 1.363 | 2.039 | 1.674\* | 1.263 | 2.829\* |
|  |  | 1.08 | 4.53 | 1.38 | 1.78 | 2.27 | 0.79 | 2.36 |
| Highest level of education: Degree or higher |  | 1.221 | 3.049\*\*\* | 1.465 | 1.650 | 1.746\* | 1.669 | 2.657\* |
|  |  | 0.74 | 5.19 | 1.54 | 1.14 | 2.31 | 1.62 | 2.13 |
| Index of Multiple Deprivation: 2nd decile |  | 0.805 | 1.122 | 1.147 | 1.039 | 0.730 | 1.870\* | 0.718 |
|  |  | 0.89 | 0.64 | 0.61 | 0.10 | -1.51 | 2.33 | -0.88 |
| Index of Multiple Deprivation: 3rd decile |  | 1.064 | 1.141 | 1.100 | 0.862 | 0.804 | 1.939\* | 0.508 |
|  |  | 0.26 | 0.68 | 0.42 | -0.39 | -1.03 | 2.37 | -1.61 |
| Index of Multiple Deprivation: 4th decile |  | 0.871 | 1.121 | 1.057 | 0.592 | 0.956 | 0.988 | 0.863 |
|  |  | 0.51 | 0.56 | 0.23 | -1.27 | -0.20 | -0.04 | -0.37 |
| Index of Multiple Deprivation: 5th decile |  | 0.916 | 1.545\* | 1.322 | 1.129 | 0.929 | 1.384 | 0.964 |
|  |  | 0.33 | 2.12 | 1.12 | 0.31 | -0.31 | 1.07 | -0.09 |
| Index of Multiple Deprivation: 6th decile |  | 0.928 | 0.963 | 0.980 | 1.135 | 0.958 | 1.428 | 0.781 |
|  |  | 0.26 | -0.17 | -0.07 | 0.32 | -0.18 | 1.12 | -0.60 |
| Index of Multiple Deprivation: 7th decile |  | 0.838 | 1.022 | 1.255 | 1.001 | 1.009 | 1.504 | 0.928 |
|  |  | 0.60 | 0.10 | 0.87 | 0.00 | 0.04 | 1.27 | -0.17 |
| Index of Multiple Deprivation: 8th decile |  | 0.971 | 1.427 | 1.219 | 1.707 | 1.525 | 2.353\*\* | 1.141 |
|  |  | 0.10 | 1.53 | 0.68 | 1.26 | 1.71 | 2.68 | 0.33 |
| Index of Multiple Deprivation: 9th decile |  | 1.018 | 1.434 | 1.210 | 0.530 | 1.091 | 1.203 | 0.703 |
|  |  | 0.06 | 1.54 | 0.68 | -1.38 | 0.34 | 0.55 | -0.83 |
| Index of Multiple Deprivation: highest decile |  | 1.128 | 1.228 | 1.275 | 0.968 | 1.106 | 0.722 | 1.029 |
|  |  | 0.42 | 0.90 | 0.85 | -0.08 | 0.40 | -0.95 | 0.07 |
| Choice made was outside 3km: Yes |  | 1.331 | 1.740\*\*\* | 1.623\* | 3.310\*\*\* | 2.486\*\*\* | 1.771\*\* | 2.552\*\*\* |
|  |  | 1.44 | 3.82 | 2.53 | 4.83 | 6.04 | 2.89 | 3.89 |
| N | 6118 |   |   |   |   |   |   |   |
| Proportion of choices the model correctly predicts | 0.585 |  |  |  |  |  |  |  |
| Proportion of choices the model correctly predicts (10% margin of error) | 0.616 |  |  |  |  |  |  |  |
| Proportion of choices the model correctly predicts: Socio-economic status: lowest quintile | 0.572 |  |  |  |  |  |  |  |
| Proportion of choices the model correctly predicts: Socio-economic status: 2nd quintile | 0.538 |  |  |  |  |  |  |  |
| Proportion of choices the model correctly predicts: Socio-economic status: 3rd quintile | 0.588 |  |  |  |  |  |  |  |
| Proportion of choices the model correctly predicts: Socio-economic status: 4th quintile | 0.546 |  |  |  |  |  |  |  |
| Proportion of choices the model correctly predicts: Socio-economic status: highest quintile | 0.560 |  |  |  |  |  |  |  |
| McFadden's R2 | 0.144 |  |  |  |  |  |  |  |
| AIC | 13601.4 |  |  |  |  |  |  |  |
| BIC | 17130.8 |  |  |  |  |  |  |  |
| Joint test for significance: SES | 0.0736 |  |  |  |  |  |  |  |
| Joint test for significance: Parent self-reported faith | 0.000 |  |  |  |  |  |  |  |
| Joint test for significance: Parent characteristics | 0.000 |  |  |  |  |  |  |  |
| Joint test for significance: Child characteristics | 0.273 |  |  |  |  |  |  |  |
| Joint test for significance: IMD | 0.154 |   |   |   |   |   |   |   |
| Note: Absolute t statistics are in the second row. Significance levels are represented by \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001. All schools in the narrow feasible choice set for each pupil were allocated to school groups based on the school's KS2 mean score, proportion of the school eligible for free school meals and faith status. They were randomly allocated between groups if the schools' characteristics were equal to the median value of FSM or KS2 in the feasible choice set. The reference school group is group 3, the 'poor, low-scoring non-faith' school. Parental preferences (represented by the 1st nominated school on the LA form) was replaced with the school that was 'liked but not applied to' by the parent. School characteristics are averaged over all schools in the 'school group', except for distance rank which is taken as the minimum of the 'school group'. School characteristics are as follows: the proportion of the school in 2005 that were eligible for free school meals, and indicator for low income households; the proportion of (final year) pupils that got a level 5 (above average) in all KS2 exams in 2003; the proportion of the school in 2005 that had some level of English as and additional language; the proportion of the school in 2005 that were White British; the proportion of the school in 2005 that had some level of special educational need; distance rank (closest, 2nd closest etc.). Variables included in the model but omitted from the table are: the main respondent’s ethnicity; faith (self-reported in wave 1 of the MCS); how often the main respondent helps the child with reading; the number of places visited with the child in the past 6 months; whether the household has regular use of a car; whether the child was the first born in the family; the number of siblings the child has; and the gender, season of birth, emotional development (constructed from the MCS) and foundation profile score (entry level assessment at primary school) of the child.  |

|  |
| --- |
| **Table 5: Conditional logit: Comparing choice set assumptions** |
|  Odds ratios | FCS-C | FCS-T |
| Proportion FSM | 0.248\* | 0.642 |
|  | 2.10 | 0.61 |
| Proportion FSM squared | 0.974 | 0.0230\*\* |
|  | 0.03 | 3.28 |
| KS2 mean (averaged over 3 years) | 11.88\*\*\* | 8.663\*\*\* |
|  | 6.40 | 4.57 |
| KS2 mean (averaged over 3 years) squared | 0.957\*\*\* | 0.962\*\*\* |
|  | 6.17 | 4.51 |
| Proportion with EAL  | 1.740 | 1.781 |
|  | 1.74 | 1.50 |
| Proportion White British | 1.170 | 2.347\* |
|  | 0.52 | 2.39 |
| Proportion with some level of SEN | 0.694 | 0.552 |
|  | 1.31 | 1.85 |
| Distance rank | 0.567\*\*\* | 0.772\*\*\* |
|  | 21.56 | 26.57 |
| Distance rank squared | 1.035\*\*\* | 1.004\*\*\* |
|  | 11.96 | 12.63 |
| N | 6118 | 6195 |
| Proportion of choices the model correctly predicts | 0.585 | 0.521 |
| Proportion of choices the model correctly predicts (10% margin of error) | 0.616 | 0.559 |
| McFadden's R2 | 0.144 | 0.161 |
| Joint test for significance: SES | 0.0736 | 0.000 |
| Joint test for significance: Parent self-reported faith | 0.000 | 0.000 |
| Joint test for significance: Parent characteristics | 0.000 | 0.000 |
| Joint test for significance: Child characteristics | 0.273 | 0.414 |
| Joint test for significance: IMD | 0.154 | 0.000 |
| Note: Absolute t statistics are in the second row. Significance levels are represented by \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001. All schools in the narrow feasible choice set for each pupil were allocated to school groups based on the school's KS2 mean score, proportion of the school eligible for free school meals and faith status. They were randomly allocated between groups if the schools' characteristics were equal to the median value of FSM or KS2 in the feasible choice set. The reference school group is group 3, the 'poor, low-scoring non-faith' school. Parental preferences (represented by the 1st nominated school on the LA form) was replaced with the school that was 'liked but not applied to' by the parent. School characteristics are averaged over all schools in the 'school group', except for distance rank which is taken as the minimum of the 'school group'. School characteristics are as follows: the proportion of the school in 2005 that were eligible for free school meals, and indicator for low income households; the proportion of (final year) pupils that got a level 5 (above average) in all KS2 exams in 2003; the proportion of the school in 2005 that had some level of English as and additional language; the proportion of the school in 2005 that were White British; the proportion of the school in 2005 that had some level of special educational need; distance rank (closest, 2nd closest etc.). Variables included in the model but omitted from the table are: the main respondent’s ethnicity; faith (self-reported in wave 1 of the MCS); how often the main respondent helps the child with reading; the number of places visited with the child in the past 6 months; whether the household has regular use of a car; whether the child was the first born in the family; the number of siblings the child has; and the gender, season of birth, emotional development (constructed from the MCS) and foundation profile score (entry level assessment at primary school) of the child.  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 6: Robustness checks** |  |  |  |  |
|  Odds ratios | (1)Full sample: FCS-C | (2)“Equal Preferences” areas | (3)Households that haven’t moved since child was born | (4)Oldest siblings in the household |
| Proportion FSM | 0.248\* | 0.0648\*\* | 0.269 | 0.0436\*\* |
|  | 2.10 | 3.26 | 1.25 | 3.10 |
| Proportion FSM squared | 0.974 | 8.924 | 0.953 | 6.320 |
|  | 0.03 | 1.60 | 0.03 | 1.13 |
| KS2 mean (averaged over 3 years) | 11.88\*\*\* | 8.787\*\*\* | 27.70\*\*\* | 18.23\*\*\* |
|  | 6.40 | 4.44 | 5.33 | 4.86 |
| KS2 mean (averaged over 3 years) squared | 0.957\*\*\* | 0.962\*\*\* | 0.945\*\*\* | 0.950\*\*\* |
|  | 6.17 | 4.35 | 5.02 | 4.68 |
| Proportion with EAL  | 1.740 | 2.409\* | 2.423 | 1.798 |
|  | 1.74 | 2.32 | 1.75 | 1.20 |
| Proportion White British | 1.170 | 1.515 | 1.480 | 1.504 |
|  | 0.52 | 1.15 | 0.80 | 0.90 |
| Proportion with some level of SEN | 0.694 | 0.396\* | 0.646 | 0.789 |
|  | 1.31 | 2.44 | 0.94 | 0.54 |
| Distance rank | 0.567\*\*\* | 0.550\*\*\* | 0.520\*\*\* | 0.554\*\*\* |
|  | 21.56 | 17.29 | 14.93 | 15.01 |
| Distance rank squared | 1.035\*\*\* | 1.039\*\*\* | 1.042\*\*\* | 1.035\*\*\* |
|  | 11.96 | 9.76 | 8.72 | 8.06 |
| N | 6118 | 4066 | 2866 | 2721 |
| Proportion of choices the model correctly predicts | 0.585 | 0.568 | 0.593 | 0.554 |
| Proportion of choices the model correctly predicts (10% margin of error) | 0.616 | 0.597 | 0.625 | 0.583 |
| McFadden's R2 | 0.144 | 0.158 | 0.168 | 0.163 |
| Joint test for significance: SES | 0.074 | 0.006 | 0.294 | 0.293 |
| Joint test for significance: Parent self-reported faith | 0.000 | 0.000 | 0.000 | 0.000 |
| Joint test for significance: Parent characteristics | 0.000 | 0.014 | 0.062 | 0.062 |
| Joint test for significance: Child characteristics | 0.273 | 0.174 | 0.668 | 0.253 |
| Joint test for significance: IMD | 0.154 | 0.340 | 0.242 | 0.530 |
| Note: See previous table. Absolute t statistics are in the second row. Significance levels are represented by \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001. All schools in the narrow feasible choice set (FCS-C). Column 2 shows the coefficients for the sample including only those that live in local authorities which operated the “equal preferences” ranking system in 2004 when the parents applied for their school place. Column 3 shows the coefficients for the sample including only those whose parents have not moved house since they were born. Column 4 shows the coefficients for the sample including only those who are the oldest sibling in the household.  |

**Table 7: Calculation of core distance-characteristic tradeoffs**

The values are the implied change in the characteristic required to compensate for the change in distance rank, in SD units

|  |  |  |
| --- | --- | --- |
|  | (1) | (2) |
|  | Distance rank1 to 2 | Distance rank3 to 4 |
|  |  |  |
| Distance – Academic quality trade-off*(Units are SDs of academic quality)* | 1.325 | 0.894 |
|  |  |
| Distance – SES Composition trade-off*(Units are SDs of SES composition)* | 2.267 | 1.733 |

Note: Uses estimates from Table 6. This is the ‘core’ trade-off, defined in equation (6), computed at average characteristics

|  |
| --- |
| **Table 8: Would representative parents in the highest and lowest socio-economic (SES) quintiles make the same choices given the same schools? The probability of choosing each school-group available for representative individuals**  |
|  | Probability that a family chooses a particular school group |
| Choice set | Representative low SES | Fictional high SES with low SES choices |
| Low SES – rich, high-scoring, non-faith school | 0.58 | 0.84 |
| Low SES – poor, low-scoring, non-faith school | 0.25 | 0.10 |
| Low SES – rich, low-scoring, faith school | 0.16 | 0.06 |
| Choice set | Representative high SES | Fictional low SES with high SES choices |
| High SES – poor, low-scoring, non-faith school | 0.18 | 0.31 |
| High SES – poor, high-scoring, non-faith school | 0.28 | 0.33 |
| High SES – rich, high-scoring, faith school | 0.53 | 0.36 |
| Note: High SES refers to parents in the highest socio-economic quintile, low SES to parents in the lowest socio-economic quintile. The representative families were found by matching their characteristics to the average for each characteristic in their quintile. For example, the representative high SES family we found has a degree as their highest level of education, and lives in the most affluent IMD decile. The representative low SES family has GCSE grades A\*-C as their highest level of education, and lives in the lowest IMD decile. All characteristics included in our main regression were averaged within the quintile and matched on. In column 3 the representative high SES individual is given the choice from the representative low SES individual’s choice set, and the representative low SES individual is given the choice from the representative high SES individual’s choice set.  |

# Appendices – Not for final publication

**Appendix Table 1: summary of variables used in analysis**

|  |  |  |  |
| --- | --- | --- | --- |
| Variable  | Proportion | Mean | Standard deviation |
| School characteristics |  |  |  |
| Proportion of the school with free school meals, 2005 |   | 0.16 | 0.15 |
| KS2 mean score for the school, 2003, 2004, 2005 |   | 27.80 | 1.49 |
| Proportion of the school with English as an additional language, 2005 |   | 0.13 | 0.22 |
| Proportion of the school that was white, 2005 |   | 0.84 | 0.25 |
| Proportion of the school that had some level of special educational need (including those with a statement of SEN), 2005 |   | 0.19 | 0.09 |
| Distance rank of school |  | 3.57 | 2.08 |
| Parental characteristics: |   |   |   |
| Socio-economic status (SES): |   |   |   |
| SES lowest quintile  | 0.188 |   |   |
| SES 2nd quintile | 0.196 |   |   |
| SES 3rd quintile | 0.199 |   |   |
| SES 4th quintile | 0.213 |   |   |
| SES highest quintile | 0.204 |   |   |
| Highest level of education: |   |   |   |
| None | 0.104 |   |   |
| 'Other' or vocational  | 0.063 |   |   |
| GCSE grades D-G  | 0.110 |   |   |
| GCSE grades A-C | 0.334 |   |   |
| AS/A level or diploma in higher ed.  | 0.193 |   |   |
| Degree + | 0.183 |   |   |
| Ethnicity: |   |   |   |
| White British  | 0.773 |   |   |
| Indian | 0.039 |   |   |
| Pakistani | 0.062 |   |   |
| Bangladeshi  | 0.024 |   |   |
| Black Caribbean  | 0.014 |   |   |
| Black African  | 0.016 |   |   |
| Other | 0.062 |   |   |
| Faith in wave 1: |   |   |   |
| None | 0.433 |   |   |
| Church of England  | 0.219 |   |   |
| Roman Catholic  | 0.088 |   |   |
| Other Christian denomination  | 0.106 |   |   |
| Muslim | 0.108 |   |   |
| Other faith  | 0.043 |   |   |
| Single parent | 0.167 |  |  |
| No regular use of car | 0.116 |  |  |
| How often help child with reading: |   |   |   |
| every day  | 0.560 |   |   |
| several times a week  | 0.319 |   |   |
| Number of places visited with child in the past 6 months |   |   |   |
| 0 | 0.024 |   |   |
| 1 | 0.067 |   |   |
| 2 | 0.148 |   |   |
| 3 | 0.230 |   |   |
| 4 | 0.275 |   |   |
| 5 | 0.208 |   |   |
| 6 | 0.048 |   |   |
| Index of Multiple Deprivation (IMD) decile: |   |   |   |
| IMD lowest decile (most deprived) | 0.155 |   |   |
| IMD 2nd decile | 0.112 |   |   |
| IMD 3rd decile | 0.111 |   |   |
| IMD 4th decile | 0.098 |   |   |
| IMD 5th decile | 0.099 |   |   |
| IMD 6th decile | 0.077 |   |   |
| IMD 7th decile | 0.080 |   |   |
| IMD 8th decile | 0.080 |   |   |
| IMD 9th decile | 0.082 |   |   |
| IMD highest decile (most affluent) | 0.106 |   |   |
| Urban/rural location |  |  |  |
| Metropolitan | 0.382 |  |  |
| Large town | 0.148 |  |  |
| Market town | 0.084 |  |  |
| Other town | 0.283 |  |  |
| First choice school was over 3km away | 0.110  |   |   |
| Child characteristics: |   |   |   |
| First born child | 0.448 |   |   |
| Number of siblings: |   |  |   |
| Only child | 0.174 |  |   |
| 1 | 0.500 |  |   |
| 2 | 0.224 |  |   |
| 3+ | 0.102 |  |   |
| Female |  0.483 |   |   |
| Season of birth: |   |   |   |
| Autumn  | 0.255 |   |   |
| Winter | 0.257 |   |   |
| Spring  | 0.249 |   |   |
| Summer | 0.239 |   |   |

**Appendix Table 2: Socio-economic status: principal component analysis**

We create a single measure of parents’ socio-economic status (SES) for our analysis instead of including many separate but correlated variables, following the method used by the IFS in recent work. We use principal component analysis (PCA) to identify the component which explains most of the variance in SES. Based on this “1st component” (which explains 42% of the variance in SES) we split the sample into quintiles. Those in the lowest SES quintile have the lowest SES, and those in the highest SES quintile have the highest.

The variables we used in PCA are as follows:

1. Whether the main respondent reported financial difficulties. This is a binary variable taken from wave 3 of the survey.
2. The highest National Statistics Socio-Economic Status (NSSEC) of the household, taken from wave 3 of the survey.
3. Housing tenure, taken from wave 3 of the survey. Whether the household lives in privately rented accommodation, social housing, or a mortgaged/owned property are entered as binary variables.
4. Logged equivalised household income. We average measures of household income over all waves, and weight for the number of people and type of person in the household using the McClemens score. Household members were given the following scores to weight by:

First adult 0.61

Spouse/partner 0.39

Other second adult 0.46

Third adult 0.42

Subsequent adults 0.36

Dependant aged 0-1 0.09

Dependant aged 2-4 0.18

Dependant aged 5-7 0.21

Dependant aged 8-10 0.23

Dependant aged 11-12 0.25

Dependant aged 13-15 0.27

Dependant aged 16+ 0.36

**Appendix Table 3: Independent School**

|  |  |
| --- | --- |
| Child attends a fee paying school | Weighted percentage (observations) |
| No | 95.46 (8,982) |
| Yes | 4.54 (399)  |

4.54% of our sample attends a fee paying school; a small minority.

|  |  |
| --- | --- |
| Child attends a fee paying school | Applied on an LA form(weighted row percentage and observations) |
|  | No | Yes |
| No | 22.47 (2,299) | 77.41 (6,670) |
| Yes | 78.63 (310) | 21.37 (89) |

21.37% of those that attend a fee paying school also applied to a state school via the LA application form.

|  |  |
| --- | --- |
| Child attends a fee paying school | Applied on an LA form but did not get 1st choice; an *ex post* move to the private sector(weighted row percentage and observations) |
|  | No | Yes |
| Yes | 37.71 (38) | 62.29 (51) |

Those that attend a fee paying school but also applied to the private sector tend to move sectors *ex post*; 62.29% of those that attend a fee-paying school but also applied to their LA did not get their first choice. This is a minority of parents compared with the total sample however.

**Appendix Table 4: Specifying School-home distance**

We have very comprehensive data about the children and parents in our sample, and are able to determine all their feasible schools. We also have very good school level data, which allows us to look at the characteristics of these feasible schools for each family. This allows us to run a comprehensive model of parents’ demand for schools.

There is one restriction in our data however; for confidentiality reasons we do not know the precise distance between the home and each school. Instead we know a distance rank (whether the school is the closest, 2nd closest, 3rd closest, and so on), which is a discrete variable. This is a problem when we want to calculate trade-offs that parents make between school characteristics. Ideally we would have a continuous variable, and would be able to ask: “if the school were an extra 500m away (for example), how much would the academic standards of the school have to increase for the probability that a family applies there to remain the same?” Instead, we can only ask: “if the school is an extra distance rank away (so 2nd closest as opposed to closest for example), how much would the academic standards of the school have to increase for the probability that a family applies there to remain the same?” As the variable is discrete we may not capture much of the subtlety of the distance measure. The variable is also a problem as the distance between schools ranked number 1 and number 2 will vary between families, for example those in rural and urban areas.

We therefore have to choose how to assign a distance rank measure to the school group and here we undertake a number of robustness checks to validate the approach we use.

For other school characteristics, we have assigned the mean value of the variables for all in the group. For the distance rank, we have tried three options:

1. Assign the school-group the mean value of distance rank of all schools in the group
2. Assign the school-group the median value of distance rank of all schools in the group
3. Assign the school-group the minimum value of distance rank of all schools in the group

The first option has the benefit that the measure is slightly more continuous. The third measure is perhaps more intuitive, however, as all other characteristics being equal, the parent is likely to choose the closest school.

We also have to choose how to enter this imperfect measure of distance in our model. There are three options that are feasible, given below:

1. Enter distance rank as a discrete variable
2. Enter distance rank as a series of binary variables
3. Enter distance rank as a continuous variable, with the addition of a squared term

If we enter distance rank as a discrete variable, we effectively impose that the effect of moving an extra school away must be constant. For example, the change in academic standards would have to be the same whether the school moved from closest to 2nd closest or from 8th to 9th closest for the probability of applying to remain constant. This may be unrealistic, as parents may be more indifferent between the 8th and 9th schools away (as there are both relatively far, and may both involve a different form of transport compared to the closest school), than between nearer schools.

If we enter the distance rank as a series of binary variables, with the closest school being the reference category, we allow the effect of distance rank to be non-linear. As expected, this model shows that parents care more about a school moving from being the closest to 2nd closest school than between distance ranks further away. We see that parents care a lot about a school moving from being their closest to 2nd closest school (as the coefficient is large and negative), but care less about moving from the 4th to the 5th school away.

The third option is to enter distance rank as a continuous variable, and also include a squared term. To make the variable as continuous as possible, we use the mean rank of all schools in the group for the school-group characteristic. This option has the benefit of simplicity, and it captures the form we think parents preferences are likely to take, i.e. that they will care more having to travel a greater distance for schools that are closer.

The graph below shows the impact of distance rank on the probability that parents choose a particular school.

As described earlier we see that option 1 forces the probability of choosing a particular school to be constant as distance rank increases. In contrast, option 2 allows the probability to vary; we see a large negative impact when the school is moved from closest to 2nd closest, and smaller impacts for further schools. The third option smoothes the two approaches. We think this best approximates the impact of distance on parents’ choice of school. We adopt this approach for our modelling in the paper however it is reassuring that the coefficient on distance rank does not vary hugely regardless of method used.

|  |
| --- |
| **Appendix Table 5: The percent of pupils with each school group in FCS-C** |
|   | % of pupils with group in choice set |
| School group | All | Low SES | High SES |
| 1. Rich, low scoring non-faith school | 35.7 | 34.3 | 37.2 |
| 2. Rich, high scoring non-faith school | 56.9 | 55.2 | 58.5 |
| 3. Poor, low scoring non-faith school | 79.1 | 86.3 | 73.6 |
| 4. Poor, high scoring non-faith school | 38.3 | 35.5 | 39.4 |
| 5. Rich, low scoring faith school | 30.4 | 26.6 | 33.0 |
| 6. Rich, high scoring faith school | 75.5 | 78.8 | 73.5 |
| 7. Poor, low scoring faith school | 45.2 | 37.6 | 53.6 |
| 8. Poor, high scoring faith school | 27.8 | 23.3 | 30.6 |
| Note: The table shows weighted percentages. The narrow feasible choice set (FCS-C) is all schools for which the pupil lives within the schools' catchment area. Catchment areas are defined by the straight line distance in which 80% of pupils in the previous cohort lived. Those with 'low SES' are those parents who are in the lowest socio-economic quintile, 'high SES' in the highest socio-economic quintile. Socio-economic quintile was computed using principal component analysis.  |
| **Appendix Table 6: % of pupils with each combination of school groups available in their narrow feasible choice set** |
|  |
| School groups available | %: whole population | %: low SES | %: high SES |
| 2, 3 | 48.9 | 49.6 | 49.4 |
| 1, 2, 3 | 15.1 | 13.9 | 16.1 |
| 1, 2, 3, 4 | 7.7 | 6.6 | 8.3 |
| 2, 3, 6 | 36.0 | 37.0 | 37.5 |
| 2, 3, 6, 7 | 18.1 | 17.1 | 22.1 |
| 1, 2, 3, 4, 6 | 5.5 | 5.0 | 6.7 |
| 1, 2, 3, 4, 6, 7 | 2.8 | 2.7 | 3.4 |
| 1, 2, 3, 4, 5, 6, 7, 8 | 0.2 | 0.0 | 0.3 |
| Note: The table shows unweighted percentages. 1 represents school group 1, 2 represents school group 2, and so on. School groups are defined relative to the local feasible choice set, or the local median. More details about this process can be found in Appendix 2. Those with 'low SES' are those parents who are in the lowest socio-economic quintile, 'high SES' in the highest socio-economic quintile. Socio-economic quintile was computed using principal component analysis; details can be found in Appendix 4. |

1. For instance, families can choose schools outside their school district (local authority). [↑](#footnote-ref-1)
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, whilst Voluntary Aided (mainly faith) schools and Foundation schools control their own admissions, as of course do private schools. [↑](#footnote-ref-2)
3. Pupils wishing to apply to a school that controls its own admissions may also complete an additional form. [↑](#footnote-ref-3)
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. [↑](#footnote-ref-4)
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. [↑](#footnote-ref-5)
6. These relate to secondary schools, but LAs will use the same system for both primary and secondary. [↑](#footnote-ref-6)
7. All these schools are therefore essentially equivalent to the characterisation of “district” schools in the US system in the mechanism design literature. [↑](#footnote-ref-7)
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. [↑](#footnote-ref-8)
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. [↑](#footnote-ref-9)
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. [↑](#footnote-ref-10)
11. ‘High’ is defined as above 30% of the ward. Proportions are based on the 1991 national census. [↑](#footnote-ref-11)
12. In wave 1 22% of our sample are Church of England, 8.5% are Roman Catholic and 10.3% are another Christian denomination. Of those in our final sample, 5.69% change faith between wave 1 and wave 3 of the survey. Around half of the people that change faith now have an affinity with the Church of England, which could be related to the often high performance of these religious schools. Strategic changes of faith by parents are often reported in the media, although in our data only 6% change faith. Also see Allen and West (2009). [↑](#footnote-ref-12)
13. 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. [↑](#footnote-ref-13)
14. Our data shows that 94.2% of parents got their first choice school. There is 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 is 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. [↑](#footnote-ref-14)
15. Note that we do not drop all pupils who have some level of emotional or behavioural difficulty. [↑](#footnote-ref-15)
16. These are the Key stage 2 exams, compulsory for all pupils in state schools in English, maths and science. These exams 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. [↑](#footnote-ref-16)
17. 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. [↑](#footnote-ref-17)
18. These split schools are typically closely linked (for example, 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 out 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. [↑](#footnote-ref-18)
19. 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](http://www.londoncouncils.gov.uk/children/briefings/PanLondoncoordinatedadmissionssystem.htm) 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. [↑](#footnote-ref-19)
20. See Annex A of West et. al. (2009). [↑](#footnote-ref-20)
21. 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. [↑](#footnote-ref-21)
22. 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. [↑](#footnote-ref-22)
23. This final restriction is simply to ensure that schools with very large catchment areas, such as boarding schools, are excluded from the feasible choice set. [↑](#footnote-ref-23)
24. Having done this, we anonymise the school district/local authority variable so we cannot identify pupils’ region of the country. [↑](#footnote-ref-24)
25. Note that the probabilities displayed here do not add to 1 for each family type. These probabilities are averages over all families and not all families have all the plotted school groups available; the probabilities do add to one within a family. [↑](#footnote-ref-25)
26. Defined as either a degree or ‘A level’ exams: the latter are academic qualifications taken at age 16-18. [↑](#footnote-ref-26)
27. This is strengthened by the fact that IMD variables remain largely insignificant in the model even when distance rank is not controlled for. This shows that the catchment area restrictions we account for are effective. When the same test is tried for the FCS-T, IMD coefficients are highly significant, which suggests that in a broader definition of FCS, location has a significant impact on parents’ choice of school. [↑](#footnote-ref-27)
28. Note that to ensure convergence of the model we re-group all faith schools into two groups: a rich, high-scoring faith school-group (as before) and an “other faith” school-group. [↑](#footnote-ref-28)
29. We explicitly note the group-specific constants because our groups are defined on school characteristics and a change in a school characteristic may switch its group. [↑](#footnote-ref-29)
30. This is then the standard way of describing the trade-offs (see Davies et al (2001)). [↑](#footnote-ref-30)
31. The mean distance for one distance rank is 800 metres. [↑](#footnote-ref-31)
32. The mean difference between first nearest and second is 812.5m; between second and third closest schools is 667.9m. The average difference between the four closest schools for each pupil is 611m. This varies substantially across the country however; pupils in Cornwall travel an extra 1977m to their second closest school, while pupils in Camden (a London LA) travel only an extra 89m on average. [↑](#footnote-ref-32)
33. The sample is not big enough to make any very strong statements about ethnic composition, and this is not the main focus of this study. [↑](#footnote-ref-33)
34. The previous Labour administration encourage the start up of so called Academy schools, whilst the Coalition government is promoting so called Free Schools. Both school types are state funded semi-autonomous schools. [↑](#footnote-ref-34)