Heterogeneous Preferences and the Efficacy of Public School Choice

by

Justine S. Hastings
Department of Economics
Yale University and NBER

Thomas J. Kane
Harvard Graduate School of Education
and NBER

Douglas O. Staiger
Department of Economics
Dartmouth College and NBER

ABSTRACT

Public school choice plans are intended to increase equity and quality in public education by offering students at underperforming schools immediate academic gains from attending a higher-achieving school and by creating demand-side pressure on underperforming schools to improve. Both potential benefits implicitly assume that all parents value academics highly and will choose schools accordingly. Using data on parental choices and lottery assignments to schools we estimate a mixed logit model of demand for schools and show that heterogeneity in choice behavior across low-and high-SES families leads to disparate gains in achievement as a result of attending a chosen school and disparate demand-side pressure for schools to improve academic performance. Our results imply that public school choice may widen rather than narrow the gap in achievement.

This paper combines results from Hastings, Kane, and Staiger (2007a) and (2007b). We thank the Charlotte-Mecklenburg School Board, Office of the Superintendent, as well as the excellent school district staff for making this project possible. We would also like to thank Joseph Altonji, Patrick Bayer, Steven Berry, David Card, Philip Haile, Carolyn Hoxby, Fabian Lange, Derek Neal, Sharon Oster, Miguel Villas-Boas, Sofia Villas-Boas, and numerous seminar participants. C. Sean Hundtofte and Jeffrey Weinstein provided outstanding research assistance. The project is funded by grants from the Institution for Social and Policy Studies at Yale University and the U.S. Department of Education.

1. Introduction

Several urban public school districts are currently experimenting with public school choice plans, and the federal No Child Left Behind Act of 2001 includes a choice provision allowing parents of children in failing schools the choice to send their children to non-failing schools outside of their neighborhood. The goal of these school choice plans is two-fold. First, school choice allows disadvantaged students the immediate opportunity to benefit academically from attending a higher-performing school. Second, school choice is intended to increase pressure on failing schools to improve, through the threat of losing students, thereby improving educational quality at under-achieving schools through demand-side pressure for higher academic performance.

Both of these potential benefits from public school choice implicitly assume that all parents value academics highly and will choose schools accordingly once residential restrictions are lifted. However, parents may have a variety of reasons for choosing schools, and differences in preferences and school choices may generate both heterogeneous short-run effects of attending a first-choice school on academic outcomes (Heckman (1997), Heckman, Smith, and Clements (1997), Heckman, Urzua, and Vytlacil (2006)) and disparate demand-side pressure for academic improvement across low-achieving and high-achieving schools (e.g., vertical differentiation in a differentiated products market, Anderson, de Palma, and Thisse (1992)). For example, if disadvantaged families place less weight on academics when choosing schools, then schools serving these families will face little pressure to improve, and children of these families may not reap academic gains from changing schools. If this is the case, then school choice may have the smallest immediate impact on the families it is intended to help, and educational outcomes and opportunities may eventually become more stratified, instead of less stratified, under school choice.

In this paper, we use a unique dataset and policy experiment from the Charlotte-Mecklenburg School Public School District (CMS) in North Carolina to examine heterogeneity in parental school choice behavior and its implications for public school choice. We use data on parents' choices for schools during the implementation of district-wide school choice in CMS in 2002 to estimate a mixed-logit demand model for schools, allowing heterogeneity to influence choice behavior, flexibly, through both observable and unobservable family characteristics. We find that the weights parents place on key school characteristics are very heterogeneous, with high-income parents of high-achieving students placing the largest weights on test scores when selecting schools. We also find negative correlations between idiosyncratic preferences for school test scores and measures of school proximity, implying that willingness to travel is highest for those who value academics the most. In addition, we show that parents of each race prefer schools where their own race is the clear majority, implying that minority parents face much larger trade-offs between academics and social preferences when choosing schools.

We then use our demand estimates to examine the implications that heterogeneity in parental choice behavior has for the ability of school choice to increase pressure on low-performing schools to improve through the threat of losing students. To do this, we simulate the demand response that would result if a school were to boost average test scores, holding all else constant. We find that demand-side pressure is largest for high-performing schools that serve very elastic families and minimal for low-performing schools that serve inelastic, disadvantaged families. These results suggest that public school choice, without additional incentive mechanisms, may lead to greater educational stratification of schools serving high- and low-income families, rather than providing a competitive tide that lifts all boats.

Finally, we examine how heterogeneity in choice behavior affects the immediate gains in test scores from attending a first-choice school for children from advantaged versus disadvantaged backgrounds. CMS used lotteries to assign students to oversubscribed schools, allowing us to identify the causal effect of attending a first-choice school on gains in own academic achievement. Theory suggests that the academic gains from attending a first-choice school will be positive for parents who place a high weight on academics but potentially negative for parents who place a low weight on academics and high weights on other school characteristics that are negatively correlated with academics. Using the mixed-logit model to estimate the weight that each parent placed on academics when choosing a school, we find that students of parents with weights at the 95th percentile experienced significant rises in End of Grade test scores of approximately 0.1 student-level standard deviations, while students of parents placing little value on academics actually experienced declines in achievement. These declines were strongest for African American families that placed little weight on academics, since they generally prefer schools with more minority students (which in CMS tend to be schools with lower academic performance). These results imply that the immediate impact of attending a first-choice school broadens (instead of narrows) the gap in performance across high- and low-SES families exercising choice.

This paper makes a number of contributions. First, it examines the potential efficacy of public school choice by estimating the underlying determinants of parents' choices. We estimate and document substantial variation in parental choice behavior using a flexible discrete choice framework. We are able to identify heterogeneity in choice parameters in our model because we have rich micro data with multiple ranked choices (Berry, Levinsohn, and Pakes (2004)), variation in choice set characteristics from large-scale redistricting with the introduction of the school choice plan, and a diverse underlying student-body population.

Second, our approach allows us to examine if school choice will provide stronger demand-side pressure for low-performing schools to improve or if it will result in greater stratification across schools serving low- and high-SES families. The prior literature has focused on estimating competitive pressure at the market level and its impact on academic achievement, using measures such as HHIs (Herfendahl-Hirschman Indices) of local districts as proxies for competition (Borland and Howson (1992), Hoxby (2000), Hanushek and Rivkin (2003); Belfield and Levin (2002) review the methodology and results in this literature). This implicitly assumes that all schools (or districts) face homogeneous demand elasticity and competitive pressure. Modeling consumer (parental) choice for differentiated products (schools) allows different types of schools to face different demand elasticities, and thus we can test how public school choice may affect demand-side incentives to improve quality across low- and high-achieving schools.

Finally, we provide an economic explanation for why certain subgroups of students benefit from exercising choice. Several recent papers have estimated the average treatment effect of winning a lottery to attend a first-choice school, abstracting away from the underlying factors that led parents to select these schools. Cullen, Jacob, and Levitt (2006) examine high-school lotteries in Chicago and find no significant average impact on test scores from attending a chosen school, leading them to conclude that measurable school inputs have little causal impact on student outcomes. Others studies find positive outcomes for some subgroups or for students applying to different types of schools, but the results vary across studies (Ballou (2007), Betts et al. (2006), Hastings, Kane and Staiger (2006)). By connecting expected treatment effects with choice behavior and the trade-offs parents face when choosing schools, we provide a framework for understanding why subgroup impacts may vary across studies and across student groups as preferences, trade-offs, and choice sets vary.

This paper proceeds in four sections. The next section provides background on the CMS school choice plan. Section 3 presents our empirical model, beginning with estimation of the mixed-logit model and then describing how we use these results to estimate the extent to which school choice placed demand-side pressure on underachieving schools and the extent to which the academic gains from attending a first-choice school varied with the importance a parent placed on academics when selecting the schools. Section 4 presents our empirical results, including a description of the data, estimates from the mixed-logit model, demand simulations, and estimates using school lotteries, to identify how attending a first choice school affects a student's academic achievement. The final section concludes with some thoughts on the implications these results have for the design of school choice.

2. The CMS School Choice Plan

Before the introduction of a school choice plan in the fall of 2002, the Charlotte-Mecklenburg Public School District (CMS) operated under a racial desegregation order for three decades. In September 2001, the U.S. Fourth Circuit Court of Appeals declared the school district "unitary" and ordered it to dismantle the race-based student assignment plan by the beginning of the next school year. In December of 2001, the school board voted to approve a new district-wide public school choice plan.

In the spring of 2002, parents were asked to submit their top three choices of school programs for each of their children. Each student was assigned a "home school" in her neighborhood, often the closest school to her, and was guaranteed a seat at this school. Magnet students were similarly guaranteed admission to continue in their current magnet programs. Admission for all other students was limited by grade-specific capacity limits set by the district. Parents could choose any school in the district. However, transportation was only provided to schools in a student's quadrant of the district (the district was split into 4 quadrants called "choice zones"). The district allowed significant increases in enrollment in many schools in the first year of the school choice program in an expressed effort to give each parent one of her top three choices. In the spring of 2002, the district received choice applications from approximately

105,000 of 110,000 students. Admission to over-subscribed schools was determined by a lottery system as described below.

Once the district was declared "unitary" and the court order requiring race-based bussing was terminated, CMS could no longer draw boundaries based on the racial composition of a neighborhood. As a result, the former school assignment zones, which often paired non-contiguous black and white neighborhoods, were dramatically redrawn. Under the choice plan, 43 percent of parcels were assigned to a different elementary grade "home school" than they were assigned to the year before under the bussing system. At the middle school and high school levels, this number was 52 and 35 percent, respectively. Therefore, the 2002-2003 home school for many students is often not the school they would have been assigned at the time they chose their residence. This dramatic change in school assignment zones, the simultaneous introduction of a sweeping school choice plan, and the assignment of students to high-demand schools by lottery provides a unique opportunity to examine the implications of parental choice behavior for achievement of disadvantaged students in a public school choice plan.

2.1 Heterogeneous Choices

Table I provides an overview of student characteristics and parents' choices for elementary and middle school students old enough to take standardized tests in reading and math, broken down by race and lunch subsidy status.[1] More detail on the data and estimation sample will be provided in Section 4. Whites and Blacks each comprised approximately 45% of the student-body population. Approximately 10 percent of white students received federal lunch subsidies, while just over 60 percent of African Americans did. Minority and lunch-subsidy recipients had, on average, lower achievement (test scores were from Spring 2002 and standardized to have a mean of zero and a standard deviation of one across all students within each grade) and came from lower-income backgrounds. Neighborhood income is measured as the median income for families living in a student's census block group of the student's same race according to the 2000 Census. In addition, minority and lunch-subsidy students had lower-

[1] Throughout the analysis, we focus on students entering grades four through eight. Estimation is limited to these grades because of the lack of test scores (either baseline or end of grade) for other grades. Students do not take End of Grade tests until grade 3, so that students entering grade 4 are the first to have a baseline test available. We do not include high school students since school choice is likely influenced by factors, such as graduation rates and athletic programs, that are not central to elementary and middle school choice, and thus would be better handled in a separate analysis.

performing "home schools" as measured by average student test scores.[2] However there was substantial variation in both own performance and local school performance within each race and income category. In particular, because CMS bused students of all races to schools in different neighborhoods, as well as "midpoint" schools between ethnic neighborhoods, students of all socio-economic backgrounds had considerable variation in the academic performance of school choices within reasonable distance from their homes. For example, Table I shows that students had a considerable range of school test scores at school choice options within six miles (approximately twice the average distance to the "home school"). School test scores range from approximately -1 to 1, so students of all socio-economic backgrounds had schools ranging from the lower to the upper quartile within reasonable proximity. Although the highest scoring school within 6 miles is, on average, larger for non-poor white families, the standard deviation is large, and the variation in school average test scores for proximate schools is substantial for all socioeconomic groups. This variation along with multiple choices will help identify the underlying determinants of parents' school choices.

Since parents were guaranteed a slot in their default school, many parents listed only one or two schools on their choice forms. In particular, parents of students who were white and ineligible for lunch subsidies were more likely to fill out only a first choice. Between 48% and 67% of parents in the remaining subgroups submitted all three choices. Note that many parents who listed their home school first listed subsequent choices as well. For instance, 65% of non-white parents whose children did not receive lunch subsidies chose their home school first, while only 51% of these parents listed only a first choice. This may be because determining the first choice necessitates determining the top three choices, making it costless to list all three choices. The availability of multiple choices from those who listed their home school first will add further choice set variation and aid in the identification of demand parameters.

Despite the proximity of high-performing schools to neighborhoods of various socio-economic backgrounds, there was little unanimity in parents' choices of schools. Even within the same elementary home school assignment for 2002-03, parents listed, on average, 10.4 different first-choice elementary schools.[3] It is clear that parents have very heterogeneous preferences

[2] School average test scores are calculated as the average of the combined standardized 2001-2002 test score for students attending the school in the 2002-2003 school year. This is the test score measure we will use in the analysis, and we will discuss alternative test score measures in Section 4.
[3] This number accounts for differences in choices listed driven by differences in the prior-year's school.

over school characteristics. Figure I shows that approximately 20% of parents chose schools that had lower test scores than the school they had guaranteed admission to, suggesting that school characteristics that were potentially negatively correlated with average test scores were the strongest determinants of choice for some families. This suggests that heterogeneous preferences may play a key role in school selection and academic outcomes in public school choice.

2.2 Lottery Assignments

The district implemented a lottery system for determining admissions to oversubscribed schools. Approximately one third of the schools in the district were oversubscribed. Under the lottery system, students choosing non-home schools were first assigned to priority groups, and student admission was then determined by a lottery number. The priority groups for district schools were arranged in lexicographic order based on the following priorities:

Priority 1: Student who had attended the school in the prior year. (Students were subdivided into 3 priority groups depending upon their grade level, with students in terminal grades—grades 5, 8, and 12—given highest priority.)

Priority 2: Parent of free-lunch eligible student applying to school where less than half the students were free-lunch eligible

Priority 3: Parent applying to a school within her choice zone

Parents listing a given school as their first choices were sorted by priority group and a randomly assigned lottery number.[4] Any slots remaining after home school students were accommodated were assigned in order of priority group and random number.[5] If a school was not filled by those who had listed it as a first choice, the lottery would repeat the process with those listing the school as a second choice, using the same priority groups as above. However, for many oversubscribed schools, the available spaces were filled up by the time the second choice priority groups came up.

---

[4] The random number was assigned by a computer using an algorithm that we verified with CMS computer programmers.

[5] Once any sibling was admitted to a school, other siblings could choose to attend the school. We dropped those who were admitted to a school because of a sibling preference.

Children of parents not assigned to one of their top choices were placed on a waiting list. About 19% of students winning the lottery to attend their parents' first-choice schools subsequently attended a different school, with 13% attending their home schools instead and another 6% attending a different school entirely, with most of these students changing address. When slots became available, students were taken off the waitlist based on their lottery numbers alone, without regard for their priority groups.

2.3 Potential for Strategic Choice

The lottery mechanism used by CMS was a "first-choice-maximizer", in which slots were first assigned to all those listing a given school as a first choice before moving to those listing the school as a second or third choice. In such a mechanism, parents with poor home school options may have an incentive to misstate their preferences – not listing their most preferred school if it had a low probability of admission (Glazerman and Meyer (1995), Abdulkadiroğlu and Sönmez (2003), Abdulkadiroğlu et al. (2006)). Instead, they may have hedged their bets by listing a less preferred option with a higher probability of admission in order to avoid being assigned to their home school. Such strategic behavior would imply that parental choices would not reflect true preference orderings for schools – to the extent that parents are *not* listing their preferred match due to strategic hedging.

However, there were a number of reasons why such strategic behavior was probably rare in the first year of the choice plan that we are studying. First, this was a new school choice plan, and parents did not know the details of how the lottery system would be operated. The handful of district officials who knew the lottery details were not allowed to communicate them to parents. Parents were never given their actual lottery numbers. The district also told parents that they would make every attempt to give each student admission to one of their chosen schools and *instructed them to list what they wanted*. Even in long-standing limited choice plans where districts instruct parents to choose strategically, there is evidence than many parents do not act strategically (Pathak and Sönmez (forthcoming)). In order to accommodate demand, the district substantially expanded capacity at popular schools. In addition, the district gave a "priority boost" to low-income students choosing to attend schools with low concentrations of low-income students. Hence, choices for top schools by students with under-performing home schools would be given top priority. This would counteract the incentive for these students to hedge their choices, as outlined above.

If there were widespread strategic behavior by parents, we would expect those with low-quality default schools to hedge their bets and list less desirable schools for which they might have a higher probability of admission. Hastings, Kane and Staiger (2007a) use the redistricting to test if parents with *exogenous* changes to the quality of their home school had lower preferences for high-quality schools, as would be predicted if parents were behaving strategically, and they find no evidence that this was the case. We discuss this evidence in more detail in Section 4 and in an appendix. Overall, we did not find evidence indicating that strategic behavior played a significant role in this first year of school choice.[6]

3. Empirical Model

3.1 Demand for Schools

Our empirical model uses the parental choices, along with data on each student and school, to estimate how parents weigh different school characteristics and how this varies in the population. We estimate a mixed logit discrete choice demand model (McFadden and Train (2000), Train (2003)). Mixed logits are multinomial logit choice models with random coefficients on product attributes. By introducing individual heterogeneity in the logit coefficients, the mixed logit model allows for flexible substitution patterns – generating credible estimates of demand elasticities. The mixed logit can approximate any random utility model, given appropriate mixing distributions and explanatory variables (Dagsvik (1994), McFadden and Train (2000)).

Our model is based on a standard random utility framework. Let U_{ij} be the expected utility of individual *i* from attending school *j*. Individual *i* chooses the school *j* that maximizes his or her utility over all possible schools in the choice set. For the first choice, the individual chooses over the set of all available schools (denoted J_i^1), so that:

$$y_{ij}^1 = 1 \text{ iff } U_{ij} > U_{ik} \; \forall k \in J_i^1$$

[6] In subsequent years of school choice, when capacities at schools were no longer changed to accommodate demand, strategy may have become more important. In the second year of choice, CMS no longer made an effort to accommodate choices by changing school capacities. Many parents received none of their three choices and expressed frustration because they had made choices without knowing the probability of admittance.

$y_{ij}^1 = 0$ otherwise.

The second and third choices (identified by y_{ij}^2 and y_{ij}^3) are made in a similar manner, except that the choice sets (denoted J_i^2 and J_i^3) exclude schools already chosen by individual *i*.

We assume that utility is a linear function of expected academic achievement, A_{ij}, and non-academic factors, such as distance from home and school racial composition, V_{ij}.

(1) $$U_{ij} = \beta_i^A A_{ij} + V_{ij}$$

Parents may form their expected utility from academic achievement on a variety of factors. Three key factors include the school's test scores, their children's own academic ability, and the families' income level.[7] Hence, the expected value of academic achievement can be written as $\beta_i^A A_{ij} = \beta_i^0 S_j + z_i^{'}\beta_i^1 S_j + z_i^{'}\beta_i^2 + v_{ij}$, where school average test scores, S_j, a student's own characteristics such as academic achievement, income level, and race, z_i, and their interactions with school test scores are the main predictors a parent uses for forming utility over academic achievement if her child attends school *j*. This specification allows academic achievement to impact utility idiosyncratically through heterogeneous coefficients as well as through interactions between school test scores, income, and student test scores. Heterogeneity in the value from expected academic achievement can be interpreted as either heterogeneous expected achievement or as heterogeneous tastes for achievement. Thus variation arises either because some parents value academic achievement more relative to non-academic factors, or because some parents are better informed about which schools promote academic achievement, or because some students' academic achievement is more influenced by the schools they attend.[8] Note that since z_i do not vary across choices, they will only impact a parent's choice through the interaction with school test scores. Non-academic characteristics also influence parental choices heterogeneously through $V_{ij} = X_{ij}^{'}\beta_i^X + \omega_{ij}$, where X_{ij} include distance from home, availability of bussing, whether the child attended the school in the prior year, whether the school was

[7] We will present results for different measures of school academic performance in the Robustness Checks appendix. These will show, for example, that parents do not use actual measures of school value-added when choosing schools but instead use school average test scores. In other robustness checks not presented in tables, we also examined if parents chose schools based on own-race average test scores and rejected that specification against the one that uses school average test scores.

[8] We estimated a handful of specific production function models and found that they performed poorly in terms of predicting the magnitude of the treatment effect when combined with the randomized lottery admissions and student outcomes, suggesting that the assumptions were too restrictive relative to the more general framework presented here.

designated as the child's home school, and the school fraction minority and its square, as well as interactions with a student's own race and income status. Substituting these assumptions into equation (1) yields the following utility model for school choice:

(2) $$U_{ij} = \beta_i^0 S_j + z_i^{'} \beta_i^1 S_j + X_{ij}^{'} \beta_i^X + \varepsilon_{ij}$$

We assume that that ε_{ij} is distributed *i.i.d.* extreme value and that the idiosyncratic portions of preferences are drawn from a multivariate normal mixing distribution ($\beta \sim f(\beta \mid \mu_\beta, \theta)$, where μ and θ denote the mean and variance parameters), yielding a traditional mixed-logit model.

Given these assumptions, the probability that individual i chooses schools (j^1, j^2, j^3) is given by:

$$P_i(j^1, j^2, j^3) = \Pr\{(U_{ij^1} > U_{ik} \forall k \in J_i^1) \cap (U_{ij^2} > U_{ik} \forall k \in J_i^2) \cap (U_{ij^3} > U_{ik} \forall k \in J_i^3)\}$$

(3) $$= \int \prod_{c=1}^{3} \frac{e^{X_{ij^c}\beta}}{\sum_{k \in J_i^c} e^{X_{ik}\beta}} f(\beta \mid \mu, \theta) d\beta$$

These probabilities form the log-likelihood function: [9]

(4) $$LL(X, \mu, \theta) = \sum_{i=1}^{N} \sum_{j=1}^{J_1} \sum_{k=1}^{J_2} \sum_{l=1}^{J_3} y_{ij}^1 y_{ik}^2 y_{il}^3 \ln(P_i(j,k,l))$$

We estimate the model separately by race and lunch subsidy status, thus allowing for heterogeneity in choice behavior through full interactions with these two key socio-economic variables. Within race and lunch subsidy status, we assume that the random parameters follow a joint normal distribution, where (1) the preference for distance is drawn from a negative lognormal distribution so that all people dislike commuting, (2) we allow the coefficient on school test scores to vary idiosyncratically but not coefficients on the interactions with baseline student achievement and family income[10], and (3) we allow for covariance between the coefficients on school test scores and measures of proximity (distance and home school), since these covariances could significantly impact demand-side pressure for schools to improve

[9] For students submitting fewer than three choices, the likelihood is modified in an obvious way to reflect only the probability of the submitted choices.

[10] In other words, within race and lunch subsidy segment, own baseline test score and family income influence the mean preference for school test scores but not the variance in idiosyncratic preferences for school test scores.

quality.[11] Since equations (2) and (3) do not have a closed form solution, simulation methods were used to generate draws of β from $f(\cdot)$ to numerically integrate over the distribution of β. Estimation was by the method of maximum simulated likelihood, using 100 draws of β from $f(\cdot)$ for each individual in the data set. The results were not sensitive to increasing the number of draws used.

3.2 Identification

Several aspects of the CMS school choice data and experiment are helpful for identifying the parameters in our demand model. First, the large scale redistricting that occurred with the introduction of school choice helps to identify values placed on distance separately from residential sorting. Residential sorting could overstate the importance of proximity and neighborhood schools if parents had previously located near to their preferred schools. However, the former school assignment zones often required parents to live far from their preferred school, and the large scale redistricting meant that many parents unexpectedly found themselves assigned to new neighborhood schools. Thus, in the first year of the choice plan, neither home school assignment nor distance to a school were strongly linked to parental preferences through prior residential sorting. In addition, multiple choices listed by those selecting their home schools first further separates preferences for school characteristics from residential sorting by simulating the unavailability of the neighborhood school.

Second, historic placement of schools for bussing in CMS provides wide variation in school characteristics for families in all socio-economic groups, dampening collinearity problems that may be present in other settings. For example, as was seen in Table I, students from all socio-economic groups had high-scoring schools within reasonable proximity. Third, approximately 95% of parents submitted choices for the choice plan. Thus we have data for nearly the entire student population—whereas most work using school choice data has been dependent on limited and potentially non-representative subgroups of students.

Fourth, the multiple ranked responses provided for each parents create variation in the choice set by effectively removing the prior chosen school from the subsequent choice set. This

[11] Allowing for a general covariance structure across all parameters led to instability in the estimated covariance terms in some specifications but did not significantly affect the remaining parameters or the substantive results that we report.

choice-set variation allows us to estimate the distribution of values parents place on school characteristics from observed substitution patterns for each individual – a stronger source of variation for identification than cross-sectional changes in the choice set based on geographic location (Train (2003), Berry, Levinsohn, and Pakes (2004)). Intuitively, when only a single (first) choice is observed for every individual, it is difficult to be sure whether an unexpected choice was the result of an unusual error term (ε_{ij}) or an unusually high weight placed by individual i, β_i, on some aspect of the choice. However, when an individual makes multiple choices that share a common attribute (e.g., high test scores) we can infer that the individual has a strong preference for that attribute, because independence of the additive error terms across choices would make observing such an event very unlikely in the absence of a strong preference.

3.3 Empirical Implications for School Choice

We use these demand estimates to examine two important implications of school choice for disadvantaged families. First, we estimate the extent to which school choice will place demand-side pressure on underachieving schools to improve through the threat of losing students. To do this we simulate the expected change in demand, Q_j, for each school were it to raise the average test scores of its students, holding all else constant. This demand response is the partial derivative of the probability that student *i*'s parents choose a school *j* as their first-choice school with respect to the average test score performance at school *j*, summed over all students $i = 1,\ldots,N$.

(5) $$\frac{\partial Q_j}{\partial S_j} = \sum_{i=1}^{N} \frac{\partial P_{ij}}{\partial S_j}$$

where the probability that parent i selects school j, P_{ij}, is given by the mixed-logit demand specification integrated over random draws from the estimated distribution of underlying utility weights.

Second, we examine the extent to which the academic gains from attending a first-choice school vary with the importance a parent placed on academics when selecting their schools. Prior work has estimated an average treatment effect of attending a first-choice school using lottery assignments across all students participating in school choice and has failed to find significant effects, leading to the conclusion that measurable school inputs appear to have no significant impact on student achievement (Cullen, Jacob, and Levitt (2006)). However, among students choosing an alternative school over their home school, the expected academic gain of a student randomized into the first-choice school is given by:[12]

(6) $E(\Delta A \mid \beta_i^A \Delta A + \Delta V > 0)$

where ΔA and ΔV denote the gain in academic achievement and in non-academic factors from attending the first-choice school over the guaranteed school option, respectively. Thus in a general framework, as β_i^A gets very large, the expected treatment effect alone determines choice and, therefore, must be positive for all students who choose an alternative school. However, for a student with low β_i^A (near zero) the expected treatment effect is ambiguous. If ΔA and ΔV are independent and ΔA is mean zero, then the expected treatment effect is zero, i.e. $E(\Delta A \mid \Delta V > 0) = 0$. If ΔA is negatively correlated with ΔV – as may be the case for some non-academic dimensions such as percent African American - then the treatment effect will be negative for students, unless their parents place sufficient weight on academic achievement to overcome the negative trade-off. That is, the parent of a minority student who wants her child to attend a school with same-race peers may give up gains in her child's academic achievement if the weight she places on academics is low. Hence, this basic framework generates the prediction that the expected treatment effect is positive for all students placing a large weight on academic achievement. Among students placing less weight on academic achievement, the expected treatment effect will depend on the tradeoffs that parents face; it could even be negative if expected academic achievement is sufficiently negatively correlated with other valued school characteristics.

We use lottery assignments to over-subscribed schools to estimate the impact of attending a first-choice school on standardized test scores, allowing the impact to vary explicitly with the estimated weight parents placed on academic achievement when selecting their schools. We analyze the subset of students choosing schools that were over-subscribed and limit our sample to students in marginal priority groups within those schools – priority groups for which lottery number alone determined initial admission. We ignore members of priority groups in which all students were either admitted or denied admission, since the assignment of lottery numbers had no impact on their admission status.

---

[12] Because the lottery was run as a "first-choice maximizer", most students who did not win the lottery for their parents' first-choice school were assigned to their home school.

More specifically, we estimate the following equation by instrumental variables:

(7) $$Y_{ij} = X_i\alpha + \gamma_1 Attended1stChoice_{ij} + \gamma_2 Attended1stChoice_{ij} * \hat{\beta}_i^A + \delta_j + \varepsilon_{ij}$$

where winning the lottery and winning the lottery interacted with $\hat{\beta}_i^A$, the weight parent *i* placed on academics, are instruments for whether the student attended her first-choice school and its interaction with $\hat{\beta}_i^A$, controlling for baseline characteristics (including $\hat{\beta}_i^A$) and lottery fixed-effects. $\hat{\beta}_i^A$ is a posterior estimate of the weight each parent placed on school tests scores, calculated from our demand model using Bayes' rule, as follows (Revelt and Train (1998), Train (2003)):

(8) $$E\left(\beta_i^A \mid y_i, X_{ij}, \mu, \theta\right) = \frac{\int \beta_i^A P\left(y_i \mid X_{ij}, \beta\right) f(\beta \mid \mu, \theta) d\beta}{P\left(y_i \mid X_{ij}, \mu, \theta\right)}$$

where y_i denotes the choices the parent made. This equation is the expected value of the weight parent *i* placed on academics given her characteristics, the choices she made, the characteristics of her choice set, and the distribution of demand parameters in the population. We calculate this posterior for each student in our randomized lottery admission group using 1000 draws from the estimated demand parameter distribution.[13]

Note that $\hat{\beta}_i^A$ incorporates all information about a parent's choices and the trade-offs she faced into one index summarizing the importance of school test scores in determining that parent's school choice. In addition, it depends only on baseline data that is independent of whether the student won the lottery, so its interaction with winning the lottery is therefore a valid instrument once one has conditioned on baseline data. Finally, note that coefficient estimates for terms involving $\hat{\beta}_i^A$ are not attenuated by the usual measurement error bias because the measurement error ($\beta_i^A - \hat{\beta}_i^A$) is uncorrelated with the posterior estimate $\hat{\beta}_i^A$ by construction (Hyslop and Imbens (2001)).

4. Data and Empirical Results

4.1 Data

[13] See Train (2003) p. 270 for Monte Carlo simulations of the accuracy of individual-level parameter estimates and the number of observed choice situations.

For this analysis, we obtained secure access to administrative data for all students in CMS for the year before and after the implementation of school choice. Throughout the analysis, we focus on students entering grades four through eight. Estimation is limited to these grades because of the lack of test scores (either baseline or outcome) for other grades. Students do not take End of Grade tests until grade 3, so that students entering grade 4 are the first to have a baseline test available. We do not include high school students since school choice is likely influenced by factors, such as graduation rates and athletic programs, that are not central to elementary and middle school choice, and thus would be better handled in a separate analysis.

For each student, we have the choice forms submitted to CMS, allowing each parent to specify up to 3 choices for her child's school. In addition to the parental choices our data contain student characteristics for the years before and after school choice, including geocoded residential location, race, gender, lunch-subsidy recipient status, and student test scores for standardized North Carolina End of Grade Exams in math and reading, and school assignment. We use these data to construct key covariates in the demand for schools, such as driving distance from each student to each school, an indicator for bussing availability, an indicator for the prior-year's school, measures of student-level income, student baseline academic achievement, school-level academic achievement, and school-level racial composition. The variables used in our model are described in detail in Table II.

The final estimation sample used to estimate the mixed-logit model included 36,887 students entering grades 4-8. The means and standard deviations of these variables across the 2.4 million school, student, and choice rank interactions available to our sample of students and schools are reported in columns 1 and 2 of Table III, and the mean and standard deviation of the average characteristic across students is reported for the 36,887 students are reported in columns 3 and 4.

4.2 Empirical Results

Table IV presents the results from the mixed logit demand estimation by race and lunch-recipient status. We report the estimates for the mean of each logit coefficient, along with the standard deviations and correlation coefficients (where appropriate) for the random parameter distributions. The standard errors for each of the estimates are reported in Appendix Table B.I.

All of the point estimates were precisely estimated and statistically different from zero at less than the 1 percent level.

The parameters determining the importance of school test scores in determining school choice are reported at the top of Table IV, followed by the parameters determining preferences over non-academic factors. Recall that within race and lunch-subsidy status, we allowed the weight that a parent placed on academic achievement to depend on the student's own baseline test score, neighborhood income, and a random intercept ($\beta_i^A = \beta_i^0 + z_i^{'}\beta^1$). Neighborhood income was demeaned, so that the intercept represents the coefficient for a student with average income and test score. The mean effect of school test scores is positive for all four demographic groups, implying that school test scores have a positive effect on the probability of choosing a school for the average student. For a student with average baseline test scores and average income, the mean effect of school scores is larger for non-white students than for white students among students not receiving lunch subsidies, both in absolute terms and relative to the marginal impact of an additional mile of travel distance.

The mean effect of school test scores is increasing in the student's own baseline achievement and neighborhood income, as indicated by the positive coefficients on the interaction of these variables with school test scores.[14] The effect of student baseline scores on the weights that parents place on school test scores is slightly larger in magnitude than the effect of income. A one standard deviation increase in the baseline test score is associated with a 0.3-0.7 increase in the mean coefficient on school test scores, while a one standard deviation increase in neighborhood income (about $25,000) is associated with a 0.35 increase in the mean coefficient on school test scores. In addition, within race, the mean coefficient on school test scores for lunch recipients, while lower than that for average non-lunch recipients, is consistent with what we would observe for a non-lunch recipient with neighborhood income that was roughly 3 standard deviations below average.

The coefficients on the interactions of income and baseline score with school scores demonstrate that the weight placed on school test scores varies considerably with observable characteristics. While differences in baseline test scores and income each generate a standard deviation in the weight placed on test scores of roughly 0.3-0.7 based on the calculations from

[14] For students who are eligible for lunch subsidies, we did not include the interaction with neighborhood income because all of these students are presumably very low income. In initial specifications, income interactions with school scores were insignificant for lunch-recipient samples.

the previous paragraph, the estimated standard deviation in the random coefficient on school test scores also ranges from 0.3 for whites not receiving lunch subsidies to 0.73 for whites receiving lunch subsidies. Hence, there is substantial unobserved heterogeneity in the weight that parents place on test scores. Looking ahead, this substantial variation across students in the weight placed on academics suggests that we may expect to see strong school choice selection on academic outcomes for some students and not for others. The fact that much of the heterogeneity in preferences is unobservable implies that the traditional approach of allowing the treatment effect to vary with observable characteristics, such as race or lunch status, may not completely capture heterogeneous preferences for academics.

The parameter estimates for the remaining coefficients indicate that parents face important trade-offs between academic and non-academic factors when choosing schools. The parameter estimates for distance and for the neighborhood (or "home") school capture the importance of proximity, either as distance traveled or as measured by inside or outside of the local neighborhood.[15] The average parent across all four segments highly valued proximity relative to school test scores, with the mean effect of a home school being roughly equivalent to six miles of travel distance, and being of the same order of magnitude as the impact of increasing school test scores by one student-level standard deviation. However, the standard deviation in the coefficient on home school is large relative to the mean and the negative correlation with the coefficient on school test scores implies that parents who place a low weight on the neighborhood school tend to place a high weight on school test scores. Hence parents who consider options outside of their home schools are more likely looking for high test scores when deciding which schools to pick. For the average parent, selecting a high-achieving school will require her to choose a school that is farther than her home school and a school that is not the home school, leading to trade-offs in utility between academic gains and proximity.

In addition to trading-off proximity for academics, African American parents must trade-off academic gains against the racial composition of peers. The coefficients on percent black and its square imply that the average African American parent prefers schools where approximately 80% of the student population is black (the peak of the mean quadratic preference for race),

[15] Hastings et al. (2007a) discuss the interpretation of the neighborhood school. They test if this coefficient represents a non-linearity in the preference for proximity or if it is potentially consistent with a default effect. They conclude that the preference for the neighborhood school is a neighborhood preference that is not generated by default behavior.

while the average white parent prefers schools where approximately 30% of the student population is black (recall that the district as a whole is approximately 45% African American). In CMS the percent black at a school is negatively correlated with average test scores (correlation is around -0.65), implying that African American parents must value academic achievement more than their white counterparts in order to induce them to choose a higher performing school that also has, on average, fewer African American students. Given the coefficients for the quadratic term in racial preferences, the loss in utility for black families is highest when percent black is low (less than 40%), which is precisely the range in which school average test scores are highest.

4.3 Robustness Checks for Demand Model

We conducted several robustness checks of the mixed logit model.[16] First, we estimated the mixed-logit model using alternative measures of a school's academic performance. Our baseline specification used 2002 average test scores of the students in the school in 2003. But we also estimated models using 2003 average test scores of the students in the school in 2003, 2002 average scores for students in each school in 2002, and a "value added" measure of each school's impact on academic achievement in the prior year (estimated as the school fixed effect from a regression of End of Grade test score on prior year test score and other student baseline characteristics, such as race, gender, and subsidized lunch status). We found that the parameter estimates across the three measures of average test scores were quite similar, but the model using our base specification had the highest likelihood value, and in that sense fit the choice data best. We found that the model using value added did not fit the observed choice data well at all. This should not be surprising, since school value added estimates were not available to parents in 2002.

Second, due to the way the lottery was run, parents may have had an incentive to misrepresent their true preferences. If they understood the allocation mechanism, a parent with an undesirable home school might want to hedge against being assigned to the home school. They would do so by picking less desirable schools than they actually prefer – trading off desirability for increased chance of being admitted. However, as discussed earlier, it was not at

[16] See Hastings, Kane and Staiger (2007a) for details. The key findings are reproduced in Appendix A for convenience.

all clear that parents had the information or experience in the first year of choice to understand how to exploit the incentives of the allocation mechanism. We tested for the presence of strategic behavior in the first year of choice by exploiting the redrawing of school boundaries. Many of those who lived in the same contiguous school assignment zone in 2001-02 were given different home school assignments in 2002-03. Hence, among those with the same school assignments who lived in the same neighborhood in 2001-02, some students experienced positive and others negative shocks to the quality of their guaranteed school. If strategy was a major component of parental choices, we would expect to see very different choice behavior for those with negative versus positive shocks to the quality of their home school. We did not, however, find significant differences in the mixed-logit parameters (particularly for the coefficients relating to school test scores) across families experiencing positive and negative shocks to their guaranteed school within the sub-sample of students who lived in 2001-2002 assignment zones that were split by redistricting. This suggests that strategy was not a major factor driving the mixed-logit parameter estimates in this first year of the public school choice plan.

Finally, we used the exogenous reassignment of nearly half of the students in the district to test the extent to which residential sorting affected our mixed-logit estimates. Residential sorting could lead us to overstate preferences for proximity if parents had already sorted to live next to the schools they preferred. To test for the potential effects of residential sorting on our estimates, we re-estimated our model for the subsample of students who were reassigned (whose school assignments under the bussing plan in 2001-2002 were different from their home school in 2002-2003). The estimated mixed-logit parameters were qualitatively and quantitatively similar in the reassigned subsample compared to those for the full sample, suggesting that endogenous residential location is not a major source of bias in this data. In addition, the similarity of the results is not too surprising if we believe our model is using the information in multiple choices to identify preferences. Recall that a substantial fraction of parents who listed their home school as their first choice also listed subsequent choices. For these parents, multiple choices simulate reassignment whether or not they were actually reassigned.

4.4 Implications for Demand-side Pressure to Improve Academic Achievement

The mixed-logit estimates can be used to simulate the degree to which demand-side pressure for academic improvement varied across low-and high-performing schools under public school choice. In order to examine the extent to which public school choice affects the demand for each school, we took each school individually, added .33 average student-level standard deviations to its mean school score, holding all else equal, and simulated the change in the number of students listing that school as a first choice.[17] The simulated change in demand gives us an idea of how responsive each school's demand curve is with respect to their own test score outcomes.

Figures II plots the change in number of students listing a school as a first choice by the school's original average score (each point in the figure is the result of a simulation for a different school) Because of the difference in size, we plot the results separately for elementary (Figure IIa) and middle (Figure IIb) schools. The demand response is quite different for schools that were originally high and low scoring. The upward sloping relationship implies that the demand response is greatest among schools that were already high scoring. This result reflects the parameter estimates in the mixed logit model. Parents with high preferences for school scores, and thus low preferences for their neighborhood schools, are sensitive to changes in school scores and are willing to consider schools over a relatively broad geography. These parents are both likely to only consider high-scoring schools for their children and are willing to change schools in response to an increase in score at another high-scoring school, even one that is located further away. These results imply that the demand-side incentives to focus on student performance are larger for higher-performing schools, since schools above a critical performance level compete intensely on academic quality for the quality-elastic segment of the population.

Figures III plots differences in average baseline test scores (in 2002) between the marginal students (those who are drawn in by the .33 average student-level standard deviation score increase) and students who previously enrolled in each school. The incentive for any school to improve its performance would be dampened if, in doing so, they were swamped by lower-performing students, who would bring down mean performance and potentially be more costly to educate. The fact that most points lie above the 45° line implies that the marginal students, on average, were higher performing than the students already enrolled. Again, this reflects the heterogeneity that was estimated in the mixed logit model, with higher performing students placing the highest weight on school test scores in choosing a school.

[17] This is approximately equivalent to a 10 point increase in the average percentile score for students attending that school.

The key features of the simulations reported in Figures II and III appear to be driven primarily by the estimated heterogeneity in preferences, rather than other details of the specification. In all the alternative specifications we have estimated that allowed for heterogeneity in preferences, we found that an increase in school test scores had a much larger effect on demand in high-scoring schools and attracted higher-performing students to the school (particularly at low-scoring schools). Eliminating unobserved heterogeneity in the parameters (estimating a conditional logit) reduced the simulated difference in demand response to higher-performing schools by roughly 15%. Eliminating preference heterogeneity through observable characteristics (income, race, lunch-subsidy status, and baseline test score) further decreased the difference in demand response across high and low-performing schools, leading to a low demand response across all schools. Thus, heterogeneity in preferences appears to be a key element in understanding the properties of parental demand for schools and their implications for student sorting and demand-side pressures for school quality in a public school choice program.

4.5 Heterogeneous Treatment Effect from Attending a 1st-choice School

In addition, heterogeneous choice behavior across low- and high-SES families may have immediate impacts on academic achievement for students exercising choice. To estimate the causal effect of attending a first-choice school and how it varies with the implicit weights parents place on academics, we analyze the subset of students choosing schools that were over-subscribed and limit our sample to students in marginal priority groups within those schools – priority groups for which lottery number alone determined initial admission.. We ignore members of priority groups in which all students were either admitted or denied admission – since the assignment of lottery numbers had no impact on their admission status. This allows us to use the random admission of students into a school, conditional on the school they chose, as an instrument for attending a first-choice school.[18]

We began with a sample of 37,115 students entering grades 4-8. Of these, 22,872 listed their guaranteed home school (n=19,669) or magnet continuation school (n=3,203) and, therefore, were not subject to randomization. Another 7,583 students were in groups sufficiently

[18] In some schools, the marginal priority group will consist of those who attended the school the year before, free-or reduced-lunch eligible students, or students from the choice zone. The marginal priority group may also be different for different grade levels in a school.

high on the priority list that they were not subject to the randomization. There were 3,065 students in marginal priority groups, described above as those priority groups within the schools where slots were allocated on the basis of a random number. Finally, there were 3,595 students in priority groups that were sufficiently low on the priority list that all members of the priority group were denied admission and placed on the waitlist.

Table V compares baseline characteristics of students in the marginal priority group to other students in the district. Overall, students in the marginal priority group appear to be fairly representative of students who chose a non-guaranteed school. The only notable differences are that students in the marginal priority group were more likely to be eligible for lunch subsidies than students in the waitlisted group (reflecting the priority given to eligible students), and they applied to schools with higher test scores than did students in the admitted group (reflecting capacity constraints at schools with higher test scores). Not surprisingly, students choosing non-guaranteed schools differed from students who chose a guaranteed school: they had home schools with lower test scores and higher proportions of minority and lunch-eligible students, and were more likely to be minority, poor, and doing poorly in school themselves. Thus, the marginal priority group should provide a reasonable estimate of the impact of attending one's first-choice school for a typical student who chose a non-guaranteed school (i.e., treatment on the treated). In addition, the final row in Table V shows the mean weight placed on academics ($\hat{\beta}_i^A$) for students in the marginal priority group versus other students in the district, calculated by equation (8). The mean $\hat{\beta}_i^A$ for students in the randomized group is very similar to that for students in the district as a whole.[19]

Table VI verifies the lottery randomization and examines the impact winning the lottery had on characteristics of the attended school in the 2002-2003 school year. For the experiment to be valid, baseline characteristics should be balanced across lottery winners and losers, differential attrition should be minimal, and winning the lottery should significantly increase the probability of attending a first-choice school. Table VI reports coefficients from regressions of the form

(9) $$Y_{ij} = X_i\alpha + \gamma_1 WonLottery_{ij} + \delta_j + \varepsilon_{ij}$$

[19] The low mean weight for students in the admitted group is consistent with the fact that these students were typically choosing lower-scoring and capacity unconstrained schools, as can be seen from the differences in sample means for the Average Combined Scores of the chosen school (row 8 in Table V).

where $WonLottery_{ij}$ is an indicator for whether student i won a lottery to attend school j, X_i is a vector of student baseline characteristics including gender, race/ethnicity, free lunch status, home school dummy variables, baseline test score, income, absences, suspensions, and grade retentions, and δ_j are school lottery fixed effects. The fixed effects, δ_j, are included for each school and grade, to account for the fact that the probabilities of winning the lottery varied across lotteries (Rouse 1998). Our estimation sample excludes 181 students who were in marginal priority groups but missing needed baseline characteristics, such as address (which was used in the choice model).

Each row of the first panel reports coefficients and standard errors for γ_1 from (9) where the baseline characteristic in each row is the dependent variable, Y_{ij} (and these baseline characteristics were not included as controls). None of the coefficients are significant, confirming that winning the lottery was independent of baseline characteristics, as would be expected if it was randomly assigned within the marginal priority groups. The second panel of Table VI tests for differential attrition, using presence in CMS at the end of the 2002-2003 school year as the dependent variable. The results show no evidence of differential attrition across lottery winners and losers.[20] The third panel of Table VI shows the reduced-form impact of winning the lottery on the characteristics of the school attended at the end of the 2002-2003 school year. Lottery winners were 53 percentage points more likely to attend the first-choice school than the lottery losers. This estimate is not equal to 100 percent for two reasons: first, some of those who were given the opportunity to attend the first choice did not do so, and second, some of those who were originally waitlisted at the first-choice school were subsequently called off the waitlist. Students who won the lottery attended schools with approximately one-tenth of a student-level standard deviation higher test scores and 7 percentage point lower lunch subsidy rates. Thus winning the lottery is a strong and valid instrument for attending a first-choice school and significantly increased the test score of the school students attend.

We estimated the impact of *attending* a first-choice school on academic outcomes using winning the lottery and its interaction with $\hat{\beta}_i^A$ as instruments for attending a first-choice school in the following specification:

[20] Average attrition rates were fairly low at 9.8% and consistent with estimates of inter-county mobility rates from the Census.

(10) $$Y_{ij} = X_i\alpha + \gamma_1 Attended1stChoice_{ij} + \gamma_2 Attended1stChoice_{ij} * \hat{\beta}_i^A + \delta_j + \varepsilon_{ij}$$

The prediction from (6) in section 3 is that $\gamma_2 > 0$ and larger in magnitude for students of parents facing significant trade-offs between academic achievement and factors that are negatively correlated with achievement. Controls include those listed for (9), but add $\hat{\beta}_i^A$ as an additional control variable. As noted earlier, since all of the information used to derive the preference weights was observed prior to randomization, $\hat{\beta}_i^A$ depends only on baseline data independent of whether the student won the lottery, and therefore, its interaction with winning the lottery is a valid instrument once one has conditioned on baseline data.

Table VII presents the results from estimating (10) by instrumental variables, using student combined average test score at the end of the 2002-2003 school year as the dependent variable. The first two columns present results for all students, with and without the interaction between attending a first-choice school and $\hat{\beta}_i^A$. The first column shows no significant average impact of attending a first-choice school on own test score, similar to average treatment effects in prior studies (Cullen, Jacob, and Levitt (2006)). The point estimate is very close to zero (-0.005) and has a large standard error (0.050). The second column shows that the weight parents placed on academics when selecting schools is a significant and positive determinant of how attending a first-choice school affects test score outcomes. The regression estimates imply that a one standard deviation increase (0.81) in the weight that an individual places on school test scores raises the treatment effect on the student's own test score by 0.066 standard deviations. For parents who placed no weight on test scores in their school choices, the coefficient on attending the first-choice school implies a negative (although not significant) treatment effect – the students' test scores fall by 0.143 standard deviations if they attend the first-choice school. These estimates imply a near zero impact (0.001 standard deviation score gain) of attending a first-choice school on test scores for an average student with a $\hat{\beta}_i^A$ of 1.78, and a large positive effect on test scores (about 0.10) for students at the 95th percentile of the $\hat{\beta}_i^A$ distribution.

A 0.1 standard deviation increase in a student's test score results is equivalent to a 3-4 percentile rank gain in test scores. Child development psychologists suggest that a 5 percentile rank gain in a student's test score translates into a significant cognitive gain in academic aptitude. However, estimates of the impact that test scores have on future earnings suggest that a 0.1 standard deviation in increase in test scores is worth \$10,000 to \$20,000 in net present value of future earnings (Kane and Staiger (2002)).

The second general prediction from our model is that the treatment effect for a student with low $\hat{\beta}_i^A$ (near zero) depends on whether parents face trade-offs – if expected academic achievement is negatively correlated with other valued school characteristics. Since the percent black at a school is negatively correlated with average test scores in CMS schools (correlation is around -0.65), the racial composition of a school is an important trade-off that many African American parents face. The last four columns of Table VII show estimates from (10) for two subgroups of parents: those for whom the estimated preferred percent minority at a school is less than 50% (primarily white students) and those for whom it is greater than 50% (primarily black students). We estimate (from the mixed logit results) that the average African American parent prefers schools where approximately 80% of the student population is black. Parents that prefer a school with a high proportion of African American students must value academic achievement more in order to induce them to choose a higher performing school that also has, on average, fewer African American students. Thus, all students whose parents have strong academic preferences (high $\hat{\beta}_i^A$) will have a positive gain in academic achievement from attending the first-choice school, but among students with weak academic preferences (low $\hat{\beta}_i^A$) we might expect a negative treatment effect among students that prefer a school with a high proportion African American. In other words, the interaction effect between $\hat{\beta}_i^A$ and attending a first-choice school should have a negative intercept and a steeper slope for students who have strong preferences for predominantly African American schools.

Posterior estimates of student-level preferences for school racial composition were calculated in the same way as the $\hat{\beta}_i^A$ s. Columns 3 and 4 show that the average treatment effect is positive for students whose parents prefer a predominantly white school, and there is no significant interaction with the weight that the parent places on test scores in their school choice. Among students of parents who prefer predominantly white schools, both relatively high- and low-$\hat{\beta}_i^A$ students experience academic gains from attending the first-choice school. In contrast, columns 5 and 6 indicate that students of parents who prefer a predominantly black school have a significant interaction between the estimated preference for academics and the treatment effect. High-$\hat{\beta}_i^A$ students experience academic gains from attending the first-choice school that are similar to students whose parents prefer a predominantly white school. In contrast, low-$\hat{\beta}_i^A$ students with parents that prefer a predominantly black school experience a negative effect on academic performance from attending the first-choice school.

These results suggest that heterogeneity in the relative importance parents place on academics when choosing schools and the trade-offs they face in their choice sets are both important determinants of the immediate gains from school choice. They imply that public school choice may have the smallest, if not negative, impact on the academic achievement of the students it is most intended to help – minority students from disadvantaged backgrounds. The results in Table VII suggest that, on average, school choice widens rather than narrows the gap in achievement between low- and high-SES families exercising choice.

Not only do measures of parental choice behavior explain heterogeneous impacts from attending a first-choice school, but they also highlight why subgroup impacts examined in prior studies vary through correlation between demographics and choice behavior. Table VIII shows estimates of the average treatment effect on student test scores in various subgroups of students defined on the basis of student demographics or characteristics of the school chosen. Prior studies have used these subgroups of students who on *a priori* grounds may have different underlying reasons for choosing schools that may be correlated with the expected treatment effect. Estimates for most of the subgroups remain insignificant. However, the estimated treatment effect is positive and significant for two of the subgroups (whites and students with above median income) and there is an apparent pattern of positive treatment effects for higher SES students and students applying to higher-scoring schools. The pattern of subgroup impacts is strongly related to the average weight ($\hat{\beta}_i^A$) that parents place on school test scores. Columns 3 and 4 of Table VIII report the mean $\hat{\beta}_i^A$ and its standard deviation, respectively, for students in each of the subgroups. There is a positive correlation between $\hat{\beta}_i^A$ and the estimated subgroup impacts across each of the subgroups. Figure IV plots the subgroup estimates from Table VIII against the mean $\hat{\beta}_i^A$ for each subgroup. The strong positive correlation between the two (correlation=0.89) suggests that differences in impacts across subgroups may be generated by differences in the underlying determinants of choice. Moreover, the large variation in $\hat{\beta}_i^A$ within each subgroup reported in Table VIII suggests that any of these simple subgroup impacts will capture only a fraction of the heterogeneity in outcomes if differences in the weights parents place on academics are driving heterogeneous treatment effects.

This evidence highlights three advantages of using demand estimates to identify heterogeneous treatment effects versus subgroup estimation based on observables, such as race and income. First, using a single index, rather than estimating differences in impacts for an arbitrary number of subgroups, increases the precision with which we can identify heterogeneous treatment effects by exploiting all of the within- and between-subgroup variation in preferences. Second, the $\hat{\beta}_i^A$ incorporate information on the choice set, distinguishing between students who pick a school because it is convenient versus students who pick it for its academics.[21] Third, the demand estimates give us an economic interpretation of subgroup impacts, allowing us to evaluate the impact of school choice outside of the estimation sample, and to potentially design school choice plans that address differences in the underlying drivers of parental choice (Hastings and Weinstein (forthcoming)).

5. Conclusion

One of the most important goals of public school choice plans is to increase academic performance for disadvantaged students by allowing them to attend higher-performing schools and by creating pressure on failing schools to improve through the threat of losing students. Both of these goals require that parents value academics and choose schools accordingly when offered the opportunity to do so. This paper departs from the prior empirical literature on public school choice by combining rich choice data, lottery assignments, and student outcomes to examine the implications that heterogeneous choice behavior has for equity and quality of educational opportunities for disadvantaged families.

[21] Note that they also include information on all parental choices and characteristics of all choice sets through the demand parameter estimates. They effectively ask how unusual a parent's choices were along the academic dimension relative to what the average person would have picked if faced with the same demographic characteristics and choice set.

We find that heterogeneity in the weights parents place on school test scores drive both demand-side pressure for schools to improve and immediate academic gains from attending a first-choice school. Because low-SES families place less emphasis on academics when choosing schools, school choice plans lead to high demand response for high-performing schools serving high-SES families and little demand response for low-performing schools serving local, low-SES families. Thus, demand response under public school choice would tend to increase education stratification rather than produce a competitive tide that lifts all boats.

Second, we find that the effect of attending a first-choice school depends on the implicit value that parents placed on test scores when choosing schools for their children. Students of parents who placed high weights on school test scores experienced gains in tests scores as a result of attending the first-choice schools. In contrast, students of parents who placed low weights on school test scores experienced (insignificant) declines in test scores as a result of attending the first-choice schools. These differences in preferences and trade-offs generate significant differences in subgroup impacts, leading to positive immediate benefit from choice to non-minority families, and negative (but insignificant) impacts for minorities. In school choice, when parents want improved academic outcomes, they are able to get them. When parents value other school attributes, and are willing to trade off academic gains for utility gains on other dimensions, school choice will allow them to make that choice – even if maximizing parental utility does not maximize academic achievement. In general, the marginal impact of school choice on academic outcomes will depend on both the willingness of parents to make these tradeoffs and the extent to which the available school choices require such tradeoffs to be made.

Finally, understanding the economic underpinnings of heterogeneous treatment effects and heterogeneous pressures to improve academics can point us towards simple changes in public school choice design that may increase efficacy in providing greater equity and quality in public education. For example, interventions such as changing the information set provided to low-income families (Hastings and Weinstein (forthcoming)) or the strategic placement of schools may increase demand-elasticity among low-SES families and minimize the trade-offs they face, allowing them to share in the potential benefits from public school choice.

References

Abdulkadiroğlu, Atila, Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez. (2006). "Changing the Boston School Choice Mechanism: Strategy-proofness as Equal Access." *Working Paper, Harvard University*.

Abdulkadiroğlu, Atila, and Tayfun Sönmez. (2003). "School Choice: A Mechanism Design Approach." *American Economic Review*, 93(3):729-747.

Altonji Joseph G., Todd E. Elder, and Christopher R. Taber. (2005). "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling." *Journal of Human Resources*, 40(4):791-821.

Anderson, Simon P., André de Palma, and Jacques-François Thisse. (1992). *Discrete Choice Theory of Product Differentiation*. Cambridge, MA: MIT Press.

Ballou, Dale. (2007). "Magnet Schools and Peers: Effects on Student Achievement." *Working Paper, Vanderbilt University*.

Belfield, Clive R., and Henry M. Levin. (2002). "The Effects of Competition Between Schools on Educational Outcomes: A Review for the United States." *Review of Educational Research*, 72(2):279-341.

Berry, Steven, James Levinsohn, and Ariel Pakes. (2004). "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market." *Journal of Political Economy*, 112(1):68-105.

Betts, Julian R., Lorien A. Rice, Andrew C. Zau, Y. Emily Tang, and Cory R. Koedel. (2006). *Does School Choice Work? Effects on Student Integration and Achievement*. San Francisco, CA: Public Policy Institute of California.

Blank, Rolf K., Robert A. Dentler, D. Catherine Baltzell, and Kent Chabotar. (1983). *Survey of Magnet Schools: Analyzing a Model for Quality Integrated Education*. Washington, DC: ABT Associates, Inc.

Borland, Melvin V., and Roy M. Howsen. (1992). "Students' Academic Achievement and the Degree of Market Concentration in Education." *Economics of Education Review*, 11(1):31-39.

Bryk, Anthony, Valerie Lee, and Petter Holland. (1993). *Catholic Schools and the Common Good*. Cambridge, MA: Harvard University Press.

Coleman, James S., Thomas Hoffer, and Sally Kilgore. (1982). *High School Achievement: Public, Catholic, and Private Schools Compared*. New York, NY: Basic Books.

Cullen, Julie Berry, Brian A. Jacob, and Steven Levitt. (2006). "The Effect of School Choice on Student Outcomes: Evidence from Randomized Lotteries." *Econometrica*, 74(5):1191-1230.

Dagsvik, John K. (1994). "Discrete and Continuous Choice, Max-Stable Processes, and Independence from Irrelevant Attributes." *Econometrica*, 62(5):1179-1205.

Evans, William N., and Robert M. Schwab. (1995). "Finishing High School and Starting College: Do Catholic Schools Make a Difference?" *Quarterly Journal of Economics*, 110(4):941-974.

Gamoran, Adam. (1996). "Student Achievement in Public Magnet, Public Comprehensive, and Private City High Schools." *Educational Evaluation and Policy Analysis*, 18(1):1-18.

Glazerman, Steven, and Robert H. Meyer. (1995). "Public School Choice in Minneapolis." In Thomas A. Downes, and William A. Testa (Eds.), *Midwest Approaches to School Reform: Proceedings of a Conference Held at the Federal Reserve Bank of Chicago October 26-27, 1994*, (pp. 110-126). Chicago, IL: Federal Reserve Bank of Chicago.

Green, Jay P., Paul E. Peterson, and Jiangtao Du. (1997). "The Effectiveness of School Choice: The Milwaukee Experiment." *Harvard University Education Policy and Governance Occasional Paper 97-1*.

Hanushek, Eric A., and Steven G. Rivkin. (2003). "Does Public School Competition Affect Teacher Quality?" in Caroline Minter Hoxby (Ed.), *The Economics of School Choice*, (pp. 23-47). Chicago, IL: University of Chicago Press.

Hastings, Justine S., and Jeffrey M. Weinstein. (forthcoming). "Information, School Choice, and Academic Achievement: Evidence from Two Experiments." *Quarterly Journal of Economics*.

Hastings, Justine S., Thomas J. Kane, and Douglas O. Staiger. (2006). "Gender and Performance: Evidence from School Assignment by Randomized Lottery," *American Economic Review Papers and Proceedings,* 96(2):232-236.

Hastings, Justine S., Thomas J. Kane, and Douglas O. Staiger. (2007a). "Parental Preferences and School Competition: Evidence from a Public School Choice Program," *NBER Working Paper No. 11805*.

Hastings, Justine S., Thomas J. Kane, and Douglas O. Staiger. (2007b). "Preferences and Heterogeneous Treatment Effects in a Public School Choice Lottery," *NBER Working Paper No. 12145*.

Heckman, James. (1997). "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used In Making Program Evaluations." *Journal of Human Resources*, 32(3):441-462.

Heckman, James J., Jeffrey Smith, and Nancy Clements. (1997). "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts." *Review of Economic Studies*, 64(4):487-535.

Heckman, James J., Sergio Urzua, and Edward Vytlacil. (2006). "Understanding Instrumental Variables in Models with Essential Heterogeneity." *Review of Economics and Statistics*, 88(3):389-432.

Hoxby, Caroline M. (2000). "Does Competition Among Public Schools Benefit Students and Taxpayers?" *American Economic Review,* 90(5):1209-1238.

Hyslop, Dean R., and Guido W. Imbens. (2001). "Bias from Classical and Other Forms of Measurement Error." *Journal of Business and Economic Statistics*, 19(4):475-481.

Kane, Thomas J., and Douglas O. Staiger. (2002). "The Promise and Pitfalls of Using Imprecise School Accountability Measures." *Journal of Economic Perspectives*, 16(4):91-114.

Krueger, Alan B., and Pei Zhu. (2004). "Another Look at the New York City School Voucher Experiment," *American Behavioral Scientist*, 47(5):658-98.

Mayer, Daniel P., Paul E. Peterson, David E. Myers, Christina Clark Tuttle, and William G. Howell. (2002). *School Choice in New York City After Three Years: An Evaluation of the School Choice Scholarships Program*. Princeton, NJ: Mathematica Policy Research, Inc.

McFadden, Daniel, and Kenneth Train. (2000). "Mixed MNL Models for Discrete Response." *Journal of Applied Econometrics*, 15(5):447-470.

Neal, Derek. (1997). "The Effects of Catholic Secondary Schooling on Educational Achievement." *Journal of Labor Economics,* 15(1):98-123.

Pathak, Parag A., and Tayfun Sönmez. (forthcoming). "Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism." *American Economic Review*.

Peterson, Paul E., David Myers, and William G. Howell. (1998). "An Evaluation of the New York City School Choice Scholarship Program: The First Year." Cambridge, MA, Mathematica Policy Research and the Harvard Program on Education Policy and Governance.

Revelt, David, and Kenneth Train. (1998). "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level." *Review of Economics and Statistics*, 80(4):647-657.

Roth, Alvin E., and Ido Erev. (1995). "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term." *Games and Economic Behavior*, 8(1):164-212.

Rouse, Cecilia. Elena. (1998). "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program." *Quarterly Journal of Economics*, 113(2):553-602.

Train, Kenneth E. (2003). *Discrete Choice Methods with Simulation*. Cambridge, United Kingdom: Cambridge University Press.

Witte, John F., Troy D. Sterr, and Christopher A. Thorn. (1995). *Fifth-Year Report: Milwaukee Parental Choice Program*. Madison, Wisconsin: University of Wisconsin.

Figure I: Distribution of Difference in Average Standardized School Score Between Student's First-Choice School and Home School

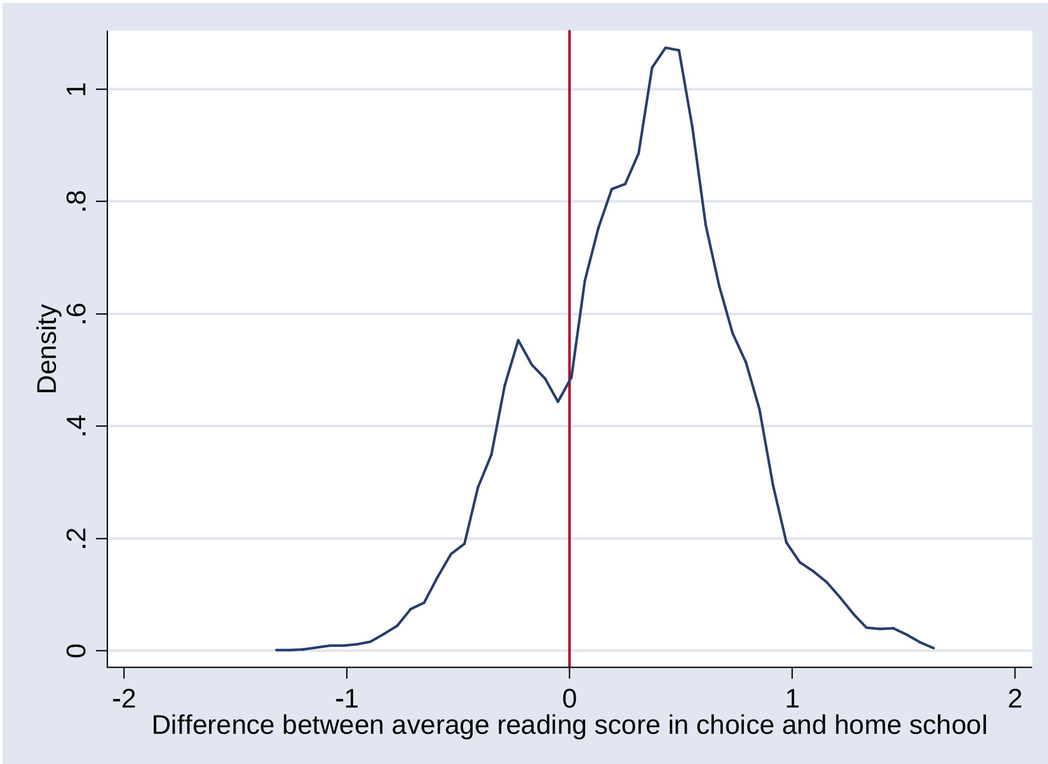

Figure IIa: Elementary Schools: Simulated Change in Number of Students Choosing School *j* When Average Standardized Score at School *j* Increases by 0.33 Points

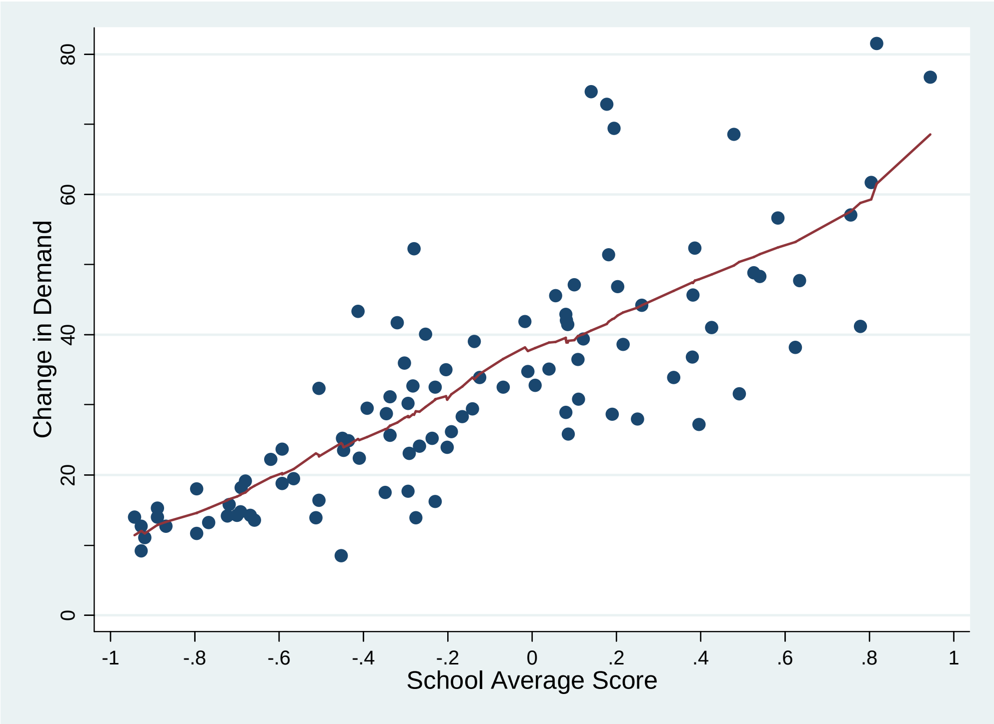

Figure IIb: Middle Schools: Simulated Change in Number of Students Choosing School *j* When Average Standardized Score at School *j* Increases by 0.33 Points

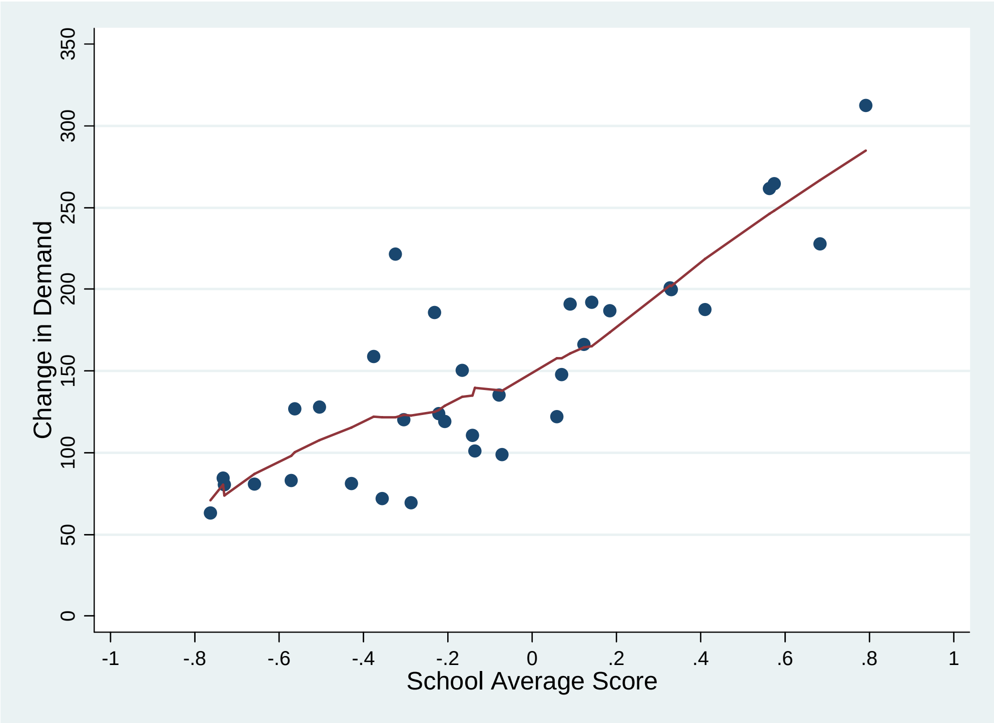

Figure III: Average 2002 Standard Deviation Scale Score for Additional Students Who Choose School *j* in Response to 0.33 Point Increase in Average Score at School *j*

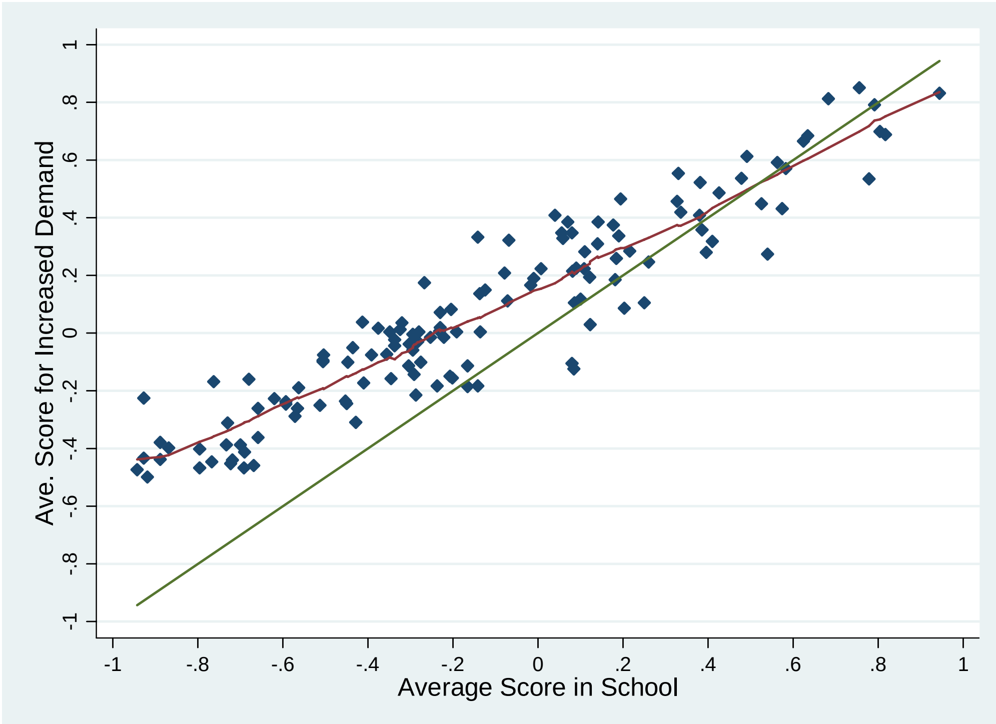

Figure IV: Subgroup Estimates of the Effect of Attending a First-Choice School

Table I: Descriptive Statistics for Students, Schools, and Parental Choices

	Not Receiving Lunch Subsidies		Receiving Lunch Subsidies	
	White	**Not White**	**White**	**Not White**
Student Characteristics				
Baseline Test Score	0.638	-0.091	-0.089	-0.613
	(0.825)	(0.839)	(0.849)	(0.799)
Neighborhood Income	$73,804	$50,612	$52,763	$36,455
	(25,869)	(21,502)	(22,556)	(16,230)
Student-School Characteristics				
Home School Ave. Test Score	0.239	-0.160	-0.147	-0.385
	(0.384)	(0.368)	(0.377)	(0.357)
Range of School Test Scores	0.670	0.817	0.827	1.018
(<6 miles)	(0.503)	(0.438)	(0.448)	(0.367)
Highest Scoring School	0.530	0.288	0.317	0.314
(<6 miles)	(0.342)	(0.317)	(0.329)	(0.268)
Choice Characteristics				
Submitted 1 Choice	0.512	0.254	0.321	0.181
Submitted 2 Choices	0.202	0.186	0.203	0.149
Submitted 3 Choices	0.286	0.559	0.475	0.670
Chose Home School	0.647	0.405	0.529	0.361
N	16,187	6,194	2,134	12,372

Notes: Standard deviations are given in parentheses. Baseline test score is the average of each student's standardized test score in reading and in math on the North Carolina End of Grade Exams. Neighborhood income is the median block group income for families of the student's own race according to the 2000 US Census. All school level test scores are reported in student-level standard deviation units. Each school's score is calculated using the average baseline combined test score (2001-2002 school year) of students who attended the school in the 2002-2003 school year. White includes Asian students who constitute less than 4% of the student population. Not White includes Hispanic and multi-racial students who constitute approximately 6% of the student population. Lunch recipients are defined as students who qualified for either free- or reduced-price federal lunch subsidies.

Table II: Explanatory Variable Definitions

Variable	Description
Distance	Driving distance from student *i* to school *j* calculated using MapInfo with Census 2000 TIGER/Line files.
School Score	Average of the 2001-2002 student-level standardized scale score for End of Grade Math and Reading exams for students in school *j* in the 2002-2003 school year. This is the average test score variable described below across all students in school *j*.
Test Score	The sum of student *i*'s scale score on End of Grade math and reading exams in the baseline year 2001-2002 standardized by the mean and standard deviation of district-wide scores for students in his or her grade.
Income	The median household income reported in the 2000 Census for households of student i's race in student i's block group. Income is demeaned by the county-wide average of approximately $51,000 and is reported in thousands of dollars.
Percent Black	The percent of students in school *j* who are black according to 2002-2003 school year administrative data.

Table III: Explanatory Variable Summary Statistics

Variable	Mean	St.Dev.	Mean of Student-level Mean	St. Dev. of Student-level Mean
Distance	13.151	6.707	12.973	3.852
Busing Provided	0.246	0.431	0.251	0.043
School Test Score	-0.115	0.453	-0.104	0.028
Student Score	0.013	0.322	0.052	0.393
Student Score* School Score	-0.007	0.463	-0.003	0.106
Neighborhood Income (demeaned, thousands of dollars)	1.096	8.97	2.367	11.623
Income * School Score	-0.679	13.025	-0.538	2.973
N	2,305,367	2,305,367	36,887	36,887

Notes: Baseline test score is the average of each student's standardized test score in reading and in math on the North Carolina End of Grade Exams. Neighborhood income is the median block group income for families of the student's own race according to the 2000 US Census. Income used in the analysis is demeaned by the district-wide mean of $51,000 and divided by 1,000. All school level test scores are reported in student-level standard deviation units. Each school's score is calculated using the average baseline combined test score (2001-2002 school year) of students who attended the school in the 2002-2003 school year.

Table IV: Estimates from Mixed Logit Model

Variable	Parameter	Not Receiving Lunch Subsidies: White	Not Receiving Lunch Subsidies: Black	Receiving Lunch Subsidies: White	Receiving Lunch Subsidies: Black
		Parameter Estimates[a]			
Preferences for Scores					
School Score	**Mean**	1.663	2.340	0.650	1.267
	Std. Dev.	0.311	0.297	0.729	0.471
Income* School Score	**Mean**	0.014	0.014	--	--
	Std. Dev.	--	--	--	--
Baseline Own Score* School Score	**Mean**	0.475	0.398	0.289	0.344
	Std. Dev.	--	--	--	--
Preferences for Proximity					
Distance[b]	**Mean**	-0.346	-0.271	-0.366	-0.281
	Std. Dev.	0.062	0.041	0.101	0.062
Home School	**Mean**	2.056	1.710	1.930	1.744
	Std. Dev.	0.848	0.910	1.508	1.586
Preferences for Race					
Percent Black	**Mean**	3.717	5.830	2.312	3.471
	Std. Dev.	2.823	1.817	2.289	1.136
Percent Black Sqaured	**Mean**	-5.259	-3.660	-3.639	-2.254
	Std. Dev.	--	--	--	--
Implied Preferred % Black	**Mean**	0.353	0.796	0.318	0.770
	Std. Dev.	0.268	0.248	0.314	0.252
Other Preferences					
Last-Year's School	**Mean**	3.877	3.585	3.579	2.959
	Std. Dev.	2.557	2.992	3.426	3.559
Choice Zone (Busing)	**Mean**	1.265	1.302	1.902	1.587
	Std. Dev.	0.884	1.221	1.494	1.208
Estimated Correlation Coefficients:					
Corr(Distance, School Score)		0.581	-0.327	0.458	-0.675
Corr(Distance, Home School)		0.104	0.001	-0.194	-0.106
Corr(School Score, Home School)		-0.173	-0.942	-0.824	-0.373

[a] All estimates are significant at the 1% level or higher.

[b] Distribution of preference on distance follows a log normal distribution.

Table V: Comparison of Student Characteristics

	All Students	Chose Guaranteed School	Chose Non-Guaranteed School		
			Admitted	Randomized	Waitlisted
Student Demographics					
Black	44.3%	34.6%	62.5%	59.7%	54.8%
Free or Reduced Lunch	39.2%	31.3%	60.3%	51.3%	34.3%
Student's Prior Year Performance					
Reading Test Score (SD Units)	0.02	0.15	-0.26	-0.09	-0.11
Math Test Score (SD Units)	0.02	0.16	-0.26	-0.12	-0.15
Choice School Characteristics					
Average Combined Scores	0.05	0.09	-0.09	0.08	0.10
Percent Free or Reduced Lunch	40.6%	38.6%	50.9%	36.6%	35.6%
Percent Minority	49.4%	46.2%	59.8%	50.0%	47.0%
Home School Characteristics					
Average Combined Scores	-0.08	0.03	-0.28	-0.23	-0.27
Percent Free or Reduced Lunch	47.0%	40.7%	59.3%	53.3%	56.0%
Percent Minority	53.6%	47.1%	65.3%	61.6%	63.8%
School Assignment					
Assigned to First Choice	85.4%	100.0%	100.0%	40.4%	0.0%
Assigned to Guaranteed School	72.5%	100.0%	0.0%	44.6%	74.5%
Posterior Weights					
Weight Placed on Academics - $\hat{\beta}_i^A$	1.79	1.89	1.51	1.74	1.82
Number of Students	37,115	22,872	7,583	3,065	3,595

Notes: Data from Charlotte-Mecklenburg Schools (CMS). Sample includes all students in grades 4-8 who applied to a regular or magnet school as their first choice for the 2002-2003 school year and were enrolled in CMS in the 2001-2002 school year. Students guaranteed placement because of siblings or in ESL are excluded.

Table VI: Baseline Characteristics by Treatment and Control Group

Variable	Mean	Adjusted Difference
Panel 1: Baseline Characteristics		
Black	0.597	0.011
		(0.022)
Free or Reduced Lunch	0.505	-0.015
		(0.012)
Median Income ($1000s) by Race	48.993	-0.700
and Block-Group in 2000 Census		(0.700)
Reading Test Score	-0.093	-0.025
		(0.031)
Math Test Score	-0.224	0.025
		(0.030)
Home School Average Combined Score	-0.122	0.003
		(0.013)
Home School Fraction Free or Reduced Lunch	0.532	0.001
		(0.007)
Home School Fraction Minority	0.614	-0.003
		(0.007)
Panel 2: Attrition		
Not Attending CMS in 2002-2003	0.098	-0.018
		(0.011)
Panel 3: School Attended in 2002-2003		
First-Choice School	0.460	0.533***
		(0.054)
Test Score of School Attended	-0.073	0.129**
		(0.040)
Number of Students		2,884

Notes: Sample limited to students in randomized priority groups with complete baseline data. Difference is between students admitted (won the lottery) and waitlisted (did not win the lottery). Each adjusted difference is from a separate regression of the given baseline characteristic on whether the student was randomly assigned to the first-choice school, controlling for lottery fixed effects. Standard errors adjust for clustering at the level of the first-choice school. Asterisks indicate significance (*=.05, **=.01, ***=.001).

Table VII: IV Estimates of Impact of Attending First-Choice School with Heterogeneous Treatment by Weight Placed on Academics in Choice Decision

Dependent Variable: ***Combined Score in Spring 2003***	**All Students** (1)	**All Students** (2)	**Students Who Prefer School Less Than 50% Black** (3)	**Students Who Prefer School Less Than 50% Black** (4)	**Students Who Prefer School at Least 50% Black** (5)	**Students Who Prefer School at Least 50% Black** (6)
Attended First-Choice School	-0.005	-0.143	0.133*	0.291	-0.041	-0.235*
	(0.050)	(0.090)	(0.064)	(0.189)	(0.058)	(0.099)
*Weight** attended First-Choice School		0.081*		-0.076		0.121**
		(0.034)		(0.071)		(0.045)
P-value for Interaction with *Weight*		0.020				0.009
Joint P-Value on Reported Coefficients	0.924	0.036	0.043	0.010	0.483	0.032
Observations	2,591	2,591	733	733	1,858	1,858

Notes: Each column in the table is from a separate IV regression. The dependent variable is a student's combined standardized test score in the spring of 2003. Each specification reports the coefficients on attending the first-choice school and its interaction with the weight that the parent places on test scores (*Weight*) in the school choice decision, using random assignment to the first-choice school and its interaction with *Weight* as instruments. All specifications control for lottery fixed effects, home school fixed effects, the baseline covariates listed in Table V, and a direct control for the student’s *Weight* estimate. Sample includes only students in the randomized priority group with complete baseline data. Standard errors adjust for clustering at the level of the first-choice school. Asterisks indicate significance (*=.05, **=.01, ***=.001).

Table VIII: Subgroup Estimates of Effect of Attending a First-Choice School

Sample	IV Estimate of Effect of Attending First-Choice School on Combined Test Score	Mean $\hat{\beta}_i^A$	Standard Deviation of $\hat{\beta}_i^A$	Number of Students
	(1)	(2)	(3)	(4)
All Students	-0.005 (0.050)	1.73	0.75	2581
Race:				
Non-White	-0.067 (0.058)	1.55	0.70	1790
White	0.172* (0.073)	2.12	0.71	791
Income:				
Below Median	-0.100 (0.058)	1.42	0.58	1601
Above Median	0.130* (0.063)	2.24	0.72	980
Free Lunch Eligibility				
Eligible	-0.061 (0.078)	1.15	0.38	1296
Not Eligible	0.070 (0.043)	2.34	0.54	1285
Baseline Test Score				
Below Average	-0.040 (0.055)	1.30	0.54	1386
Above Average	0.066 (0.064)	2.25	0.63	1195
First-Choice School Combined Score				
Below Median	-0.036 (0.080)	1.51	0.67	1337
Above Median	0.047 (0.043)	1.98	0.76	1244

Notes: Each row reports estimates for a different student subgroup, as indicated. Column 1 reports IV estimates of the impact of attending the first-choice school on the combined student test score, using random assignment to the first-choice school as an instrument. Regressions control for lottery fixed effects, home school fixed effects, and the baseline covariates listed in Table V. Sample includes only students in the randomized priority group with complete baseline data. Standard errors adjust for clustering at the level of the first-choice school. Asterisks indicate significance (*=.05, **=.01, ***=.001). Column 2 reports the average weight that parents place on test scores (*Weight*) in their school choice decision, calculated according to equation (8). The second column reports the standard deviation of *Weight* among families in each of the subgroup categories.

Appendix A: Robustness Checks

There are a number of reasons that the estimates in Table IV may not accurately represent student preferences. In this section we present a series of robustness checks that address three particularly important concerns that could lead our estimates to understate the strength of preferences for academic quality at a school. These specification checks do not find evidence that the preference estimates in Table IV are biased by any of these three important concerns. This Appendix is taken from the Robustness Section in Hastings, Kane and Staiger (2007a). Please see that paper for further details.

Alternative Measures of Academic Quality

In Table IV, we used average base-year (2002) test scores among students attending the school in 2003 as a measure of academic performance that influences parents' choices. However, if this is a crude or incorrect proxy for the information available to parents, then our estimates may understate the extent to which parents care about academic performance in choosing a school. Table A.I presents mixed logit results for our original measure plus three alternative measures of school academic performance. To preserve space, we only report the coefficients on the measure of school academic performance (and its interactions with income and student baseline score), along with the log likelihood for each model as an indicator of overall fit. The first panel replicates our baseline specification. This specification assumes that parents correctly forecasted which students would choose each school and used these students' scores from the prior year as an indicator of how good the school would be in terms of academics. The second panel uses the average score in 2003 of the students in the school in 2003. This measure implies that parents correctly foresaw student sorting and outcomes and made their choices based on that. The third panel uses the 2002 average scores for students in each school in 2002. This measure implies that parents used historical student assignment and outcomes as the best indicator of future school performance. The final panel uses a "value-added" measure of each school's impact on academic achievement in the prior year. We estimated each school's "value-added" by regressing a student's test score performance in 2003 on math and reading performance in the prior year, demographic characteristics, grade fixed effects, and fixed effects for each school. The fixed effect estimated for each school represents our estimate of a school's average impact on student performance.

The results in Table A.I show that the preference estimates across the three measures of average test scores are quite similar. The specification in panel 1, using our base specification, has the highest likelihood value, and in this sense it fits the choice data best. The value-added measure in panel 4 does not fit the observed choice data well at all. This may not be surprising since this statistic is not available to parents and is not easy to calculate given the available average score statistics. Cullen, Jacob, and Levitt (2006) report that high-demand schools in Chicago's high school choice program tend to be schools with high average test scores, not high value-added. Note that our specifications using test score levels do allow parents to make a crude race-adjusted "value-added" calculation when choosing a school (since these specifications separately control for the racial makeup of the school).

Overall, the results from Table A.I suggest that our preference estimates are not particularly sensitive to the use of reasonable alternative measures of school academic performance. Because our baseline specification provides the best fit of the data, we will rely on this specification for the final simulations.

Strategy

As noted above, parents may have had an incentive to misrepresent their true preferences. If they understood the allocation mechanism, a parent with an undesirable home school might want to hedge against being assigned to the home school. They would do so by picking less desirable schools than they actually prefer – trading off desirability for an increased chance of being admitted. This strategy could make it appear that low-income families (with lower-performing home schools) under-value academics even though they do not. They pick lower performing schools in order to increase their chance of admission, not because they place a lower weight on academics. However, it is not at all clear that parents had the information or experience in the first year of choice to understand how to exploit the incentives of the allocation mechanism. Parents did not know their lottery numbers or the assignment mechanism. In addition, parents were instructed by the district to list the schools they wanted on their choice form.[1]

We test for the presence of strategic behavior in the first year of choice by exploiting the redrawing of school boundaries. Many of those who lived in the same contiguous school assignment zone in 2001-02, were given different school assignments in 2002-03. Hence, among those with the same school assignments who lived in the same neighborhood in 2001-02, some students experienced positive or negative shocks to the quality of their guaranteed school. Table A.II shows the average difference for students who had positive versus negative shocks to their home school quality given that they lived in a 2001-2002 contiguous assignment zone that was split into new assignment zones. The table shows that the difference in scores was large and significant for many of the students affected by redistricting. If strategy was a major component of parental choices, we would expect to see very different choices for those with negative versus positive shocks to the quality of their home school, given a contiguous 2001-2002 school assignment zone. In particular, we should see significantly lower weight placed on average test scores for students who had a negative shock to the average quality of their home schools.

Table A.III presents a simple conditional logit specification using the first choices for the subsample of students who lived in 2001-2002 assignment zones that were split by redistricting. The results are estimated on the sample of those students who had a positive shock and those that had a negative shock to their home school quality. Standard errors are clustered at the prior-year's school assignment zone. The results show that there is no significant difference in the school-score preference estimates for redistricting losers versus winners.[2] In addition, the point estimates do not follow the expected pattern if strategy was a key component in choosing schools. Preferences for test scores are higher (in bold) as often as they are lower for the redistricting losers.

[1] This stands in contrast with other long-standing and limited choice programs, such as the Boston Public School choice program which told parents to consider carefully what schools they chose to list (Abdulkadiroglu et al. (2006)). Abdulkadiroglu et Al. (2006) also show that many parents appear to behave non-strategically in Boston Public Schools limited and long-standing school choice program and that these parents would benefit most from strategizing on their first choice by picking a less popular school first.

[2] We also tested for differences in preferences using a reduced form regression of the average test score of the chosen school on the average test score at the home school, controlling for 2001-2002 school assignment fixed effects and student demographic information. This compares the average scores of first-choice schools within 2001-2002 assignment zones across students with positive and negative shocks to home school quality. We find no significant difference in the average scores of schools chosen across redistricting winners and losers.

Overall, we do not see evidence that parents with poorer home school assignments hedged their bets. It is possible that such strategic behavior may develop over time as parents became more familiar with the system.[3] Our findings suggests that there is little evidence of hedging behavior in the first year of choice when parents most likely did not understand the allocation mechanism.

Residential Sorting

The exogenous reassignment of nearly half of the students in the district is also useful for testing the extent to which residential sorting affects our preference estimates. Residential sorting may lead us to overstate preferences for proximity if parents had already sorted to live next to the schools they prefer. What we interpret as a strong preference for proximity influencing school choice may actually be the opposite – strong preference for a school influencing proximity. Both redistricting and the multiple choices in our data will help identify preferences for proximity from preferences for other school characteristics.

To test for the potential effects of residential sorting on our estimates, we re-estimate our model for the subsample of students who were reassigned (whose school assignments under the bussing plan in 2001-2002 were different from their home school in 2002-2003).[4] Table A.IV provides summary statistics comparing the reassigned sample to the sample of students who were not reassigned. Because of the nature of the prior system of bussing, students who were reassigned were much more likely to be non-white and eligible for lunch subsidies. (The bussing plan often assigned students living in neighborhoods with large concentrations of minority students to attend school in neighborhoods with lower concentrations.) But within the four demographic groups, the reassigned students looked similar to those who were not reassigned in terms of baseline test scores and median income. More interestingly, parents of reassigned students were much less likely to choose their home school or their last year's school, and much more likely to list three choices. These facts are not necessarily evidence that parents of reassigned students have less preference for their home school: Parents of children not

[3] Evidence from laboratory experiments using simple extensive-form games between small numbers of players indicates that it takes time for players to learn how to play the game. In games with incomplete information on others' payoffs, learning and convergence to the perfect equilibrium is slower and sometimes does not occur (Roth and Erev (1995)).

[4] These students include both students whose 2001-2002 school assignment zone was split into two new home school zones as well as those whose entire 2001-2002 school assignment zone was reassigned to a new home school.

reassigned were more likely to have their home school be their last year's school, making it very likely that they would choose that school. In contrast, parents of reassigned children faced a less clear choice since their last year's school was no longer their home school.

Table A.V reports results from the mixed logit model estimated on the sample of students who were reassigned. The most striking feature of these estimates is their similarity to estimates from the full sample. Estimates of the mean and standard deviation of all the preference parameters are qualitatively and quantitatively similar. The mean of the parameter for home school is actually higher for all demographic groups in the reassigned sample, while the means for the distance and choice zone parameters are about equally likely to increase as decrease in the reassigned sample.

Overall, these estimates suggest that endogenous residential location is not a major source of bias in this data. In addition, the similarity of the results is not too surprising if we believe our model is using the information in multiple choices to identify preferences. Recall from Table I that a substantial fraction of parents who listed their home school as their first choice also listed subsequent choices. For these parents, multiple choices simulate reassignment whether or not they were actually reassigned.

Other Robustness Checks

A range of alternative specifications yielded similar quantitative and qualitative results. We have pooled elementary and middle school students for simplicity, but estimating the model separately for elementary and middle schools yielded similar parameter estimates. As already mentioned, we experimented with alternative specifications for the racial composition of the school, including dummy variables and splines in percent black. The spline estimates were very consistent with the more parsimonious quadratic specification.

We also specified distance to each school in terms of driving time (based on expected speed on each class of road) rather than driving distance, yielding nearly identical results. In addition, estimations using splines in distance indicated that the linear functional form used in our model was appropriate. We experimented with a range of alternative proxies for academic quality of a school. Using closely related measures, such as the average percentile score, resulted in nearly identical estimates. Allowing for non-linearities in the effect of school scores, through a quadratic or spline term, did not change the qualitative implications of the parameter estimates.

However these models fit the data poorly in the tails of the distribution, and for this mechanical reason they generated implausible results when used in simulations. In addition, estimates using splines in school test scores indicated that the linear model fit the data well for most segments of the population. Including separate terms for the school average test scores of whites and non-whites separately resulted in all students, both white and non-white, placing similar weights on the two scores, with both racial groups placing a larger weight on white test score performance. Again, the implications of the results were unchanged across these specifications. Finally, including a separate dummy variable for schools that were academic magnets (e.g., International Baccalaureate, Math and Science magnets) reduced the mean coefficient on school test scores by about half. This result highlights that average test scores are a proxy for the academic focus of a school and not necessarily the sole causal factor driving demand.

Finally, when we estimated a general mixed logit model with full covariance terms for the parameters, we found that some covariance terms became unstable in some specifications. For example, when we included a covariance between racial preferences and preferences for other characteristics could often be unstable, yielding corner solutions in some circumstances. However, the means and standard deviations of the preference parameters were largely unchanged, and the implications of the estimates in the demand simulations were very similar. This suggests that some of the covariance terms are poorly identified, but these terms are not of first order importance to simulations of demand.

Table A.I: Comparing Alternative Measures of School Academic Achievement

		Not Receiving Lunch Subsidies		Receiving Lunch Subsidies	
	Preference Parameter	*White*	*Black*	*White*	*Black*
Score Measure: Spring 2002 Scores of Students in School in Spring 2003					
- Log Likelihood		41311.12	32954.29	9241.04	80209.13
School Score	**Mean**	1.6627	2.3404	0.6504	1.2669
	St. Dev.	0.2530	0.2810	0.6478	0.3474
Test Score* School Score	**Mean**	0.4752	0.3981	0.2888	0.3438
	St. Dev.	--	--	--	--
Income* School Score	**Mean**	0.0139	0.0139	--	--
	St. Dev.	--	--	--	--
Score Measure: Spring 2003 Scores of Student in School in Spring 2003					
- Log Likelihood		41458.57	33201.53	9246.13	80402.94
School Score	**Mean**	1.3855	1.9640	0.5656	1.0616
	St. Dev.	0.4259	0.2421	0.6985	0.3867
Test Score* School Score	**Mean**	0.4852	0.4415	0.2763	0.3233
	St. Dev.	--	--	--	--
Income* School Score	**Mean**	0.0136	0.0119	--	--
	St. Dev.	--	--	--	--
Score Measure: Spring 2002 Scores of Student in School in Spring 2002					
- Log Likelihood		41809.18	33758.23	9306.37	80777.08
School Score	**Mean**	0.9107	1.0208	0.4068	0.4837
	St. Dev.	0.5060	0.0914	0.6163	0.4471
Test Score* School Score	**Mean**	0.4824	0.4586	0.2641	0.3057
	St. Dev.	--	--	--	--
Income* School Score	**Mean**	0.0096	0.0085	--	--
	St. Dev.	--	--	--	--
Value Added: Average Regression Adjusted Gains in Test Scores from 2002-2003[a]					
- Log Likelihood		42452.06	34120.71	9286.52	80858.73
School Score	**Mean**	-1.0668	-0.5114	-0.7711	-0.7896
	St. Dev.	1.4638	0.5670	0.9681	0.8189
Test Score* School Score	**Mean**	1.5659	1.3402	0.2098	0.2062
	St. Dev.	--	--	--	--
Income* School Score	**Mean**	0.0161	-0.0047	--	--
	St. Dev.	--	--	--	--

[a] Value Added calculated as school fixed effects in a regression of 2003 standardized test scores on 2002 standardized scores, controlling for student characteristics, such as race, lunch recipient status, and grade level. Empirical Bayes measures of Value Added were calculated and were correlated with Value Added ($\rho = 0.95$).

Table A.II: Differences in Scores Across Redistricted Polygons

		Average Difference in 2002-2003 Home School Scores Within a Given 2001-2002 School Assignment Polygon	
		Mean	**St. Dev.**
Not Receiving Lunch Subsidies	**White**	0.2925	0.3165
	Not White	0.2570	0.2552
Receiving Lunch Subsidies	**White**	0.2891	0.2836
	Not White	0.2866	0.3111

Table A.III: Seemingly Unrelated Regression Conditional Logit Estimates of Preferences for Academics Across Redistricting Winners and Losers

			Lost with Redistricting	Won with Redistricting
White	***Not Receiving Lunch Subsidies***			
		School Score	**2.5224**	2.2739
			(0.4593)	(0.3498)
		Test Score* **School Score**	0.1345	**0.4714**
			(0.1210)	(0.1388)
		Income* **School Score**	-0.0022	**0.0137**
			(0.0080)	(0.0115)
		Joint Test P-Value:		0.23
Not-White	***Not Receiving Lunch Subsidies***			
		School Score	2.6227	**3.0614**
			(0.1449)	(0.1470)
		Test Score* **School Score**	0.3625	**0.4712**
			(0.1051)	(0.1009)
		Income* **School Score**	0.0144	**0.0205**
			(0.0053)	(0.0045)
		Joint Test P-Value:		0.15
White	***Receiving Lunch Subsidies***			
		School Score	0.8037	**1.4041**
			(0.3939)	(0.3720)
		Test Score* **School Score**	**0.7460**	0.1472
			(0.2027)	(0.2408)
		Joint Test P-Value:		0.13
Not-White	***Receiving Lunch Subsidies***			
		School Score	**1.4594**	1.3982
			(0.1510)	(0.1993)
		Test Score* **School Score**	**0.4731**	0.4189
			(0.0858)	(0.0849)
		Joint Test P-Value:		0.83

Table A.IV: Summary Statistics Comparing Reassigned and Non-Reassigned Students

	Overall		No Lunch Subsidies				Lunch Subsidies			
			White		Black		White		Black	
	Non-Reass.	Reass.	Non-Reass.	Reass.	Non-Reass.	Reass.	Non-Reass.	Reass.	Non-Reass.	Reass.
%White, Non-Lunch	0.5419	0.2892								
%White, Lunch	0.0674	0.0461								
%Non-White, Non-Lunch	0.1569	0.1903								
%Non-White, Lunch	0.2338	0.4744								
Median Income	61,311	48,267	73,325	71,564	52,132	48,523	41,770	33,647	41,770	33,647
Average Z-Score	0.2215	-0.1870	0.6486	0.5747	-0.0344	-0.1540	-0.5144	-0.6720	-0.5144	-0.6720
Percent Chose Home 1st	0.6921	0.3105	0.7693	0.4070	0.5913	0.2667	0.5687	0.2734	0.5687	0.2734
Percent Chose Last Year School	0.6609	0.4042	0.7196	0.4692	0.6141	0.4264	0.5557	0.3507	0.5557	0.3507
Percent Made 3 Choices	0.3872	0.5814	0.2551	0.3727	0.5100	0.6057	0.6025	0.7019	0.6025	0.7019
Score of New Home School	0.0119	-0.2410	0.2065	0.1224	-0.1511	-0.2552	-0.2884	-0.4557	-0.2884	-0.4557
Score of Old Home School	0.0119	-0.1675	0.2065	-0.0778	-0.1511	-0.2459	-0.2884	-0.1827	-0.2884	-0.1827
Average Score Difference: Old-New	0.0000	-0.0733	0.0000	0.2025	0.0000	-0.0115	0.0000	-0.2713	0.0000	-0.2713

Table A.V: Mixed Logit Estimates for Reassigned Sub-Sample of Students

		Parameter Estimates[a]			
		No Lunch Subsidies		Lunch Subsidies	
Variable	***Preference Parameter***	***White***	***Black***	***White***	***Black***
Distance[b]	**Mean (lognormal)**	-0.3763	-0.2745	-0.3061	-0.2801
	Std. Dev. (lognormal)	0.0661	0.0408	0.0395	0.0589
Last-Year's School	**Mean**	3.5422	3.4251	3.5870	3.0111
	Std. Dev.	2.3087	2.8449	2.8059	3.4437
Home School	**Mean**	2.3071	1.8390	2.0496	1.8338
	Std. Dev.	0.8174	1.1331	1.5804	1.5533
Choice Zone	**Mean**	1.1425	1.3133	1.7984	1.6011
	Std. Dev.	0.7629	1.0372	1.2424	0.9999
School Score	**Mean**	1.5587	2.4388	0.3785	1.1470
	Std. Dev.	0.4442	0.5068	0.4040	0.6688
Test Score* School Score	**Mean**	0.5371	0.4270	0.0928	0.3071
	Std. Dev.	--	--	--	--
Income* School Score	**Mean**	0.0196	0.0197	--	--
	Std. Dev.	--	--	--	--
Percent Black	**Mean**	3.5729	5.5543	0.9015	2.3149
	Std. Dev.	2.8769	1.8569	1.8622	0.7953
Percent Black Squared	**Mean**	-5.2151	-3.4869	-2.5284	-1.4146
	Std. Dev.	--	--	--	--
Implied Preferred % Black	**Mean**	0.3426	0.7965	0.1783	0.8182
	Std. Dev.	0.2758	0.2663	0.3683	0.2811
Estimated Correlation Coefficients:					
Corr(Distance, School Score)		0.3951	-0.3571	0.5111	-0.6070
Corr(Distance, Home School)		0.1344	-0.0063	-0.1847	-0.1083
Corr(School Score, Home School)		-0.6297	-0.7740	-0.7727	-0.4230

[a] All estimates are significant at the 1% level or higher.
[b] Distribution of preference on distance follows a log-normal distribution.

Appendix B: Standard Errors for Demand Parameter Estimates

Table B.I: Standard Errors for Parameter Estimates in Table IV

	No Lunch Subsidies		Lunch Subsidies	
	White	**Black**	**White**	**Black**
Standard Errors on Mean Preferences				
Distance	0.0002485	0.0000020	0.0000055	0.0000021
Last-Year'S School	0.0003436	0.0000015	0.0000059	0.0000291
Home School	0.0002006	0.0000394	0.0000194	0.0000280
Choice Zone	0.0006687	0.0000019	0.0000040	0.0000211
School Score	0.0012939	0.0000888	0.0000192	0.0000704
Test Score* School Score	0.0004424	0.0000018	0.0000190	0.0000543
Income* School Score	0.0000521	0.0000003	--	--
Percent Black	0.0357668	0.0002074	0.0003098	0.0007694
Percent Black Squared	0.0378734	0.0000378	0.0002692	0.0006274
Standard Errors on Standard Deviations				
Distance	0.0000359	0.0000093	0.0000049	0.0000005
Last-Year's School	0.0001151	0.0001092	0.0000413	0.0000178
Home School	0.0004402	0.0000702	0.0000112	0.0000530
Choice Zone	0.0023968	0.0000583	0.0000124	0.0000459
School Score	0.0008831	0.0000249	0.0000140	0.0000272
Percent Black	0.0084486	0.0000289	0.0001043	0.0002395
Standard Errors on Correlation Coefficients				
Corr(Distance, School Score)	0.0000613	0.0003647	0.0000121	0.0002664
Corr(Distance, Home School)	0.0000817	0.0001029	0.0000046	0.0000073
Corr(School Score, Home School)	0.0051998	0.0000999	0.0000125	0.0001219