Does Competition Among Public Schools Benefit Students and Taxpayers? Comment

Jesse Rothstein*
Princeton University

February 2004

* Industrial Relations Section, Firestone Library, Princeton, NJ 08544; jrothst@princeton.edu. Phone (609) 258-4045; fax (609) 258-2907. I am grateful to Orley Ashenfelter, David Card, Alan Krueger, Robert McMillan, Cecilia Rouse, Miguel Urquiola, and participants at the UC Berkeley Public Finance seminar for helpful suggestions and discussions, though of course none are responsible for the outcome. This research would not have been possible without access to confidential data from the National Center for Education Statistics and the College Board. Financial support from an NSF Graduate Research Fellowship is gratefully acknowledged. Any errors, opinions, findings, conclusions, or recommendations expressed in this publication are mine alone and do not necessarily reflect the views of the NSF.

Does Competition Among Public Schools Benefit Students and Taxpayers? Comment

1. Introduction

Economically-minded observers have long argued that market discipline would create incentives for more efficient educational production (e.g. Friedman 1962; Brennan and Buchanan 1980; Chubb and Moe 1990). Although this argument is theoretically compelling, it is difficult to support empirically. One important difficulty is that the claim is about general equilibrium dynamics of competition among providers, and cannot be tested with small-scale interventions like the school voucher programs studied by Rouse (1998), Krueger and Zhu (2003), and many others. Although Chile and New Zealand have each adopted large-scale choice policies (Hsieh and Urquiola 2003; Fiske and Ladd 2000), there is as yet no large-scale, long-run choice implementation in the United States with which to study the effects of competition among schools.

The most compelling domestic evidence on productivity effects comes from studies of "Tiebout choice," the use of the residential location decision to choose among neighboring local monopoly providers. In an influential paper in this *Review*, Hoxby (2000) argues that metropolitan housing markets with more fragmented school governance offer more choice to consumers, which produces greater competitive pressures on providers. Measures of fragmentation, however, are potentially endogenous to school productivity. There has been a great deal of consolidation of school districts in the United States since World War II, and the extent of consolidation in any particular area may have depended on the productivity of local schools. Additionally, families with children may concentrate in desirable districts, implying that district enrollments themselves are endogenous. Hoxby proposes variation in topography—which may have influenced optimal jurisdictional size before modern transportation technologies—as a source of exogenous variation.

Instrumenting a measure of Tiebout choice in a metropolitan area with the number of streams flowing through that area, she finds substantial positive effects of choice on student test scores and other academic outcomes.[1]

This paper presents a re-analysis of this result, using to the extent possible the same data and specifications used by Hoxby. The decision to revisit Hoxby's analysis is motivated by two considerations. First, it has been proposed that estimates of choice's productivity effects are biased—most likely upward—by the use, as in Hoxby's study, of a sample of students enrolled in public schools (Hsieh and Urquiola 2003). Second, recent estimates of specifications similar to Hoxby's on alternative data sets have come to substantively different conclusions (Rothstein 2003). It is thus important to understand whether Hoxby's results are sensitive to the exact sample and specification used.

With the restricted-access National Educational Longitudinal Survey (NELS) data that Hoxby uses in her preferred specifications, and following her description of her data construction, I am unable to precisely replicate Hoxby's sample. When I estimate a specification similar to hers—I have access to only one of her two "streams" instruments, though it is by far the strongest in her first stage—on my replication sample, I obtain quite different results. My point estimates indicate a smaller, insignificant effect of choice on students' test scores. Moreover, when I use techniques that are robust to the data's likely clustered error structure, I estimate standard errors that are substantially larger than are Hoxby's, enough so that even her point estimates are indistinguishable from zero.

After presenting the replication estimates, I consider the impact on the estimated choice effect of small alterations to the sample and specification. I propose a new and substantially more

[1] Belfield and Levin (2002) review other studies of the effect of competition on student outcomes and conclude—based, in part, on the Hoxby result—that the evidence supports the claim of positive effects.

powerful instrument—capturing the same initial conditions proxied by the streams variables—for the Tiebout choice measure; I explore two techniques for ridding estimates of potential bias from selection into the private sector; I present several estimates of the sampling variance of the choice coefficient; and I extend the analysis to a new data set with observations on over 300,000 SAT-takers, drawn from nearly every high school in each sample metropolitan area, from the same cohort surveyed in the NELS.

Estimates of models similar to Hoxby's appear to be sensitive to the choice of instruments and to the exact sample definition. Although standard errors are quite large, I find some (statistically insignificant) evidence supporting the contention that estimates from public school samples are biased upward by endogenous selection into private schools. This suggests that analyses similar to Hoxby's, even if robust and precise, are limited in their ability to estimate the effects of competition on school productivity.

I conclude that Hoxby's positive estimated effect of interdistrict competition on school output is not robust, and that a fair read of the NELS evidence does not support claims of a large or significant effect. I do not find compelling evidence of endogeneity of the choice index to school quality, suggesting that the more precise negative (but insignificant) OLS estimate of choice's effect on student outcomes should be preferred to less precise IV estimates. An implementation of a specification similar to Hoxby's in the SAT-taker data supports my conclusions from the NELS.

Hoxby's empirical specification, the data, and the sample construction are described in the next two sections. Section 4 presents the first stage analysis, using as instruments for choice both a version of Hoxby's "smaller streams" variable and an alternative variable characterizing available choice in the metropolitan area in 1942. Basic replication estimates for the NELS test score models are in Section 5, as are estimates using the new instrument. Section 6 explores the robustness of these estimates to slight modifications in the sample. Section 7 presents two attempts to account for the selection bias introduced by the use of a sample of public school students. Section 8 considers the sampling variability of the OLS and IV coefficients, comparing Hoxby's "random effects" standard error estimator (Moulton 1986) with less parametric estimators. Section 9 describes the SAT data and presents estimates derived from them.

2. Empirical specification

Hoxby estimates models of the form

(1) $$t_{isdm} = \alpha + c_m \beta + X_m \gamma_m + X_{dm} \gamma_{dm} + X_{isdm} \gamma_{isdm} + X_{isdm} \gamma_{isdm} + \varepsilon_{isdm},$$

where t_{isdm} is the test score of student *i* at school *s* in district *d* in metropolitan area *m*. c_m is the choice index, a size-adjusted count of the number of districts serving MA *m*. X_m, X_{dm}, X_{sdm}, and X_{isdm} are control variables measured at the MA, district, school, and individual level, respectively. (To the extent that the same measures appear at several levels, X_{sdm} is the average of X_{isdm} among all students at school *s*, and so on, although in practice this may not hold in the data because we observe only a sample of students at school *s*, and because data at different levels derive from different sources.)

The error term in (1), ε_{isdm}, is not, in general, independent across observations in the same school, and we might write

(2) $$\varepsilon_{isdm} = e_m + e_{dm} + e_{sdm} + e_{isdm}$$

to reflect error components that are common to all observations in the same MA, district, or school.

Hoxby argues that the choice index may be endogenous to the quality of schools in the MA, i.e. that $E[c_m e_m] \neq 0$. She instruments for c_m with counts of larger and smaller streams in the metropolitan area, which I denote Z_m.

Private enrollment and selection bias

Hoxby's NELS estimates, and most of the replication estimates considered here, derive from samples consisting only of students in public schools. This can introduce selection biases. In particular, even with an exogenous instrument ($E[\varepsilon_{isdm} | Z_m] = 0$), the choice effect $\hat{\beta}_{iv}$ is biased relative to choice's effect on public school productivity if $E[e_{isdm} | s \text{ is a public school}]$ varies with choice.[2] Hoxby (Table 5) presents evidence that high-choice markets have lower private enrollment rates, suggesting that choice does indeed affect the private school selection process.

The direction and magnitude of the resulting bias depends on the residuals of the marginal students: If the students brought into the public sector by expansions of choice have lower e_{isdm}'s than do average public school students, $\hat{\beta}$ is biased downward relative to the effect of choice on public school productivity.[3] More plausibly, if those parents whose sectoral choice is influenced by the structure of local governance are above average in their involvement with their children's education, marginal students are likely to have unusually high e_{isdm}'s. Hoxby seems to make this claim when she discusses the consequence of "families with a strong taste for education leav[ing] the public sector by shifting their children into private schools" (p. 1233). In this case, the selection bias is positive, and the effect of choice on average public school scores is upward-biased relative to its effect on public school quality.[4] I present estimates in Section 6 that attempt to gauge the

---

[2] It may also be that $e_{dm}+e_{sdm}$ varies between sectors, although since we are interested in the effect of choice on public school quality it may reasonably be imposed that $E[e_{dm} + e_{sdm} | s \text{ is a public school}] = 0$, thus loading any across-market variation in average public school quality onto e_m.

[3] The discussion assumes that (1) is correctly specified to absorb all peer effects in educational production in the coefficient vector γ_{sdm}, and that there are no interactions between, for example, individual- and school-level characteristics. If (1) is not correctly specified, the biases in estimates of β are difficult to sign; these biases are exacerbated by the likely effects of choice on the stratification of students across schools (see, e.g., Rothstein 2003).

[4] From an earlier version of Hoxby's paper: "Since concentration appears to affect the sorting of students into private schools, potential for sample selection bias exists when we look only at the students who remain in public schools. If, for instance, one of the effects of greater public school enrollment concentration is making abler students more likely to

importance of this form of bias, first by adding private school students to the sample and second by including an inverse Mills ratio derived from the metropolitan private enrollment rate as a control variable in (1).

Sampling variance

Because the error structure (2) is non-classical, traditional standard errors overstate the precision of estimates. Under common assumptions, the true variance of the OLS estimator is

$$\text{(3)} \quad \begin{gathered} \text{var}\left(\hat{\beta}_{ols}\right) = \left(X'X\right)^{-1} X'\Sigma X\left(X'X\right)^{-1}, \text{ where} \\ \Sigma \equiv \text{var}\left(e_{isdm}\right) I_N + \text{var}\left(e_{sdm}\right) Q_{sdm} + \text{var}\left(e_{dm}\right) Q_{dm} + \text{var}\left(e_m\right) Q_m; \end{gathered}$$

I_N is an N-by-N identity matrix; and Q_{sdm}, Q_{dm}, and Q_m are block-diagonal matrices consisting of blocks of ones within each school, district, and metropolitan area, respectively, and zeros elsewhere. (That is, $Q_m = WW'$, where W is an N-by-M matrix of indicators for M metropolitan areas, and Q_{dm} and Q_{sdm} are defined similarly.) Moulton (1986) proposes a feasible version of (3) that replaces the variance terms with estimates from the data.[5] Moulton's approach is easily extended to the IV estimator, with:

$$\text{(4)} \quad \text{var}\left(\hat{\beta}_{iv}\right) = \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1} X'Z(Z'Z)^{-1}Z'\Sigma Z(Z'Z)^{-1}Z'X\left[X'Z(Z'Z)^{-1}Z'X\right]^{-1}.$$

Hoxby reports (p. 1219) using Moulton's formula with metropolitan- and district-level error components, thus implicitly imposing $\text{var}\left(e_{sdm}\right) = 0$. For most of the current analysis, I report both classical standard errors, which ignore the clustered structure of the data, and "clustered" standard errors, which are consistent in the presence of arbitrary within-MSA error covariances. The

---

attend private schools, then public school students' performance will be worse in concentrated SMSAs simply because selection into the public school student sample is negatively correlated with ability," (Hoxby 1994b, p. 24).

[5] FGLS or maximum likelihood estimators would be more efficient than classical OLS or IV in the presence of correlated errors. I follow Hoxby in considering only the classical OLS and IV estimators.

clustered standard errors are consistently larger than are those that Hoxby reports. I explore this issue in Section 7, where I implement several versions of the Moulton standard error estimator but continue to estimate standard errors 10-50% larger than are Hoxby's.[6]

3. Data

Hoxby estimates versions of model (1) on two data sets, the NELS and the National Longitudinal Study of Youth (NLSY). Her preferred estimates are from the NELS data, and I focus on these exclusively. The NELS reports test scores in four subjects for each student from each of three waves, the first conducted when the students were in the 8th grade in 1988 and the others (for students remaining in school) in 1990 and 1992. Hoxby presents estimates for reading scores from the 8th and 12th grade surveys, and for mathematics scores from the 10th grade survey. Below, I present replication estimates for the 8th and 12th grade reading scores; unreported estimates for other scores are similar. In my opinion, the 8th grade analyses are most compelling, both because this wave provides the largest samples and because non-random mobility and attrition may bias estimates from later waves. Thus, when I consider modifications of the basic specification, I present estimates only for 8th grade reading scores.

I attempt to define control variables similar to those used by Hoxby. Like her, I draw district-level demographic characteristics from the School District Data Book (SDDB), a tabulation of data from the 1990 Census along school district boundaries. However, where Hoxby appears to derive metropolitan area controls from the City and County Data Book (CCDB)—which reports

---

[6] There are several available estimators of these variance components, and neither Moulton nor Hoxby specifies which is to be used. My implementation estimates them from the contrast between individual and group-mean residual variances, extending Greene's (2000, pp. 570-2) discussion of random effects in unbalanced panels to the multi-level hierarchy considered here. Other consistent estimators might use the residual variance of within- and between-group versions of (1); the within-group empirical covariances; or an optimal minimum distance estimator using the entire empirical covariance matrix $\hat{\varepsilon}\hat{\varepsilon}'$.

1980 demographic characteristics—I use instead the 1990 Census Summary Tape File (STF) 3A for this purpose.[7] The metropolitan area definitions used at all points in my analysis are the Office of Management and Budget's Metropolitan Statistical Area (MSA) definitions of June 30, 1990, used to characterize metropolitan areas in 1990 census data. Each enumerated sub-area (PMSA) within the largest urban agglomerations is treated as a distinct MSA. I construct Hoxby's Herfindahl-based index of choice among school districts using enrollment data from the 1990 Common Core of Data (CCD), an annual census of public schools and districts.[8]

A major difficulty in replicating Hoxby's sample is the matching of NELS schools to MSAs, as the NELS offers several indirect indications of schools' locations but no direct MSA code. Hoxby reports an 8th grade sample size of 10,790 students, from 211 MSAs, but does not report her geocoding algorithm. I am unable to replicate her precise sample size. I come closest with a relatively crude match, using district codes from the NELS to match public schools to the 1987-88 CCD and relying on unaltered MSA codes from the CCD. Doing so, and retaining all students with valid 8th grade reading scores and background variables at schools with valid MSA codes and complete SDDB records, produces a sample of 11,495 students in 196 MSAs.[9] I use this as my base replication sample for analyses of 8th grade scores, and construct a 12th grade replication sample (5,865 students from 212 MSAs, as compared with Hoxby's 6,119 students from 209 MSAs) by a similar algorithm.

---

[7] I rely on the 1990 data because they seem to provide a more appropriate measure of demographic characteristics relevant to students who were in the 8th grade in 1988, and because the CCDB does not seem to tabulate ancestry, as would be needed to compute Hoxby's "ethnic homogeneity index" control.

[8] In principle, one would want to differentiate between elementary, secondary, and unified school districts by constructing the choice index only over enrollment in a particular grade or range of grades (Urquiola 1999). Hoxby does not discuss this issue, and I come closest to the indices reported in an earlier version of her paper (Hoxby 1994b) when I pool enrollment at all levels. In practice, indices are sensitive to the precise definition in only a small number of MSAs—the correlation between the pooled index and that for smaller grade ranges is always greater than 0.97—and I obtain similar regression results with other definitions.

[9] A somewhat different MSA count can be obtained by using different MSA definitions—the 1983 vintage, e.g., or the alternative county-based NECMAs in New England—or different waves of the CCD. I have not been able to obtain a count greater than 200 for the base year schools.

In Section 6, I present evidence on the sensitivity of the estimated choice effect to several refinements in the NELS geocoding. First, some errors in the CCD MSA codes can be repaired by careful inspection, as when the CCD reports an obsolete code for 3 NELS 8th grade schools' MSAs.[10] Another 19 schools' codes can be cleaned using CCD variables indicating the city, county, and state in which each district is located. Finally, NELS variables characterizing the demographic composition of the school's zip code provide an independent source of information—and the only source for private schools—about school locations. The zip code variables are drawn from the 1990 Census STF-3B tabulation, and matching to these data assigns nearly every NELS school to a unique zip code. Most zip codes lie either entirely within a single MSA or entirely outside any MSA, so the zip code uniquely assigns the vast majority of schools in the NELS sample. This leads to changes in 3 public schools' MSA codes, and provides MSA assignments for 174 private schools.

4. First Stage

Table 1 reports Hoxby's "first stage" model (from her Table 2) and analogous models derived from the replication sample.[11] Hoxby uses two instruments for the educational market structure, counts of "larger" and "smaller" streams flowing through each MA. Only the latter, derived from the public-use Geographic Names Information System (GNIS), is available for the current analysis.[12] In Hoxby's model, reproduced in Column 1 of Table 1, the "smaller streams"

---

[10] One example, from a non-NELS school, is the 827040 code, which indicated the St. Louis PMSA under the 1983-vintage MSA definitions. In 1990, this area is demoted to the St. Louis MSA, number 7040.

[11] The models in Table 1 are estimated on the universe of MSAs, and are therefore slightly different than the actual first stages to the IV regressions shown later, which are estimated on individual student observations from a subset of MSAs. Moreover, I follow Hoxby in including the metropolitan population shares below age 19 and above age 65 as regressors in Table 1, although they are not included in the IV models discussed below. As in the student-level sample, my MSA-level sample size differs from Hoxby's. One possible explanation is that Hoxby may be reporting the first stage from her NLSY sample, for which an earlier set of metropolitan area definitions would be appropriate.

[12] Hoxby has declined my request for use of her "larger streams" variable, which derives from her hand count from paper maps. For her smaller streams variable, Hoxby uses the longitude and latitude of GNIS streams' origin and

instrument accounts for a much larger share of the variance of the choice index than does the "larger streams" instrument. Moreover, Hoxby writes that "one has more a priori confidence in the exogeneity of the smaller streams variable because smaller streams are too small to affect modern life" (p. 1230). Thus, estimates that rely solely on the smaller streams instrument should be consistent and nearly as efficient as those from Hoxby's two-instrument specification. In any case, the replication estimate of the smaller streams coefficient in Column 2 is quite similar to the corresponding coefficient in Hoxby's model.[13]

Columns 3 and 4 report estimates using a new instrument. To the extent that the initial conditions influencing current market structure are exogenous to current school quality, as must be assumed for validity of the streams instruments, historical measures of market structure at a far enough remove also provide exogenous variation in current choice. I construct an estimate of the choice index in 1942, using counts of the independent school districts serving each county in the United States at that time (Gray 1944) and assuming that all districts were equally sized.[14] This measure is a substantially more powerful predictor of 1990 choice than are the streams variables.

5. Replication estimates

Table 2 reports basic replication estimates of OLS and IV models for the NELS 12th grade reading scores. Columns A and B report Hoxby's reported coefficient and standard error on the district choice index, estimated by OLS and IV respectively. The latter is the only model for which

---

destination to assign them to metropolitan areas; I use instead a GNIS variable that codes each county through which each stream flows.

[13] One major point of divergence is the population coefficient. Hoxby reports that her population measure is scaled in thousands, so I multiply her coefficient by 10,000 to obtain the effect-per-ten million reported in Table 1. It seems clear that Hoxby's coefficient is actually scaled similarly to mine.

[14] This is the earliest year for which I have been able to obtain the county-specific district counts needed for construction of the index. I have found one earlier observation, 1932, for the nation as a whole; nearly 85% of the post-1932 decline in the number of districts occurred after 1942, and the rate of decline was considerable greater in the 1942-1970 period than it was before. There has been essentially no net consolidation since 1970.

Hoxby reports control variable coefficients, which are useful for gauging the accuracy of the replication. Columns C and D report OLS and IV estimates of the same model using the replication sample and the smaller streams instrument. The control variable coefficients in Column D are broadly similar to those reported by Hoxby. I report both conventional standard errors, which do not account for potential serial correlation in the data, and, for the choice coefficient, standard errors adjusted for within-MSA clustering (in square brackets). Note that the clustered standard errors in Columns C and D are substantially larger than are Hoxby's reported error-components estimates in the corresponding models in Columns A and B. I return to this issue in Section 7.

The choice effect point estimates in columns C and D are notably smaller (closer to zero) than are Hoxby's estimates from corresponding models. The divergence of OLS estimates suggests that this is partially due to differences in the sample and in control variables rather than to the absence of the "larger streams" instrument in column D. However, the difference between OLS and IV estimates is much smaller in the replication sample than in Hoxby's results.

Column E replaces the smaller streams instrument with the 1942 choice index, while Column F includes both instruments simultaneously. In each case, I estimate a small, insignificant choice effect. The final row of the table reports *p* values for tests of the exogeneity of the choice index, assuming independence of observations within MSAs. Two of the three exogeneity tests are marginally significant, though unreported tests that allow for clustering do not approach significance.[15]

---

[15] A traditional Hausman test is inappropriate when disturbances are non-independent, as in Hoxby's random effects model. The reported *p*-value is from an *F*-test on the first-stage fitted choice index coefficient in an augmented version of the OLS specification (Davidson and MacKinnon 1993), which remains appropriate when OLS is not efficient even under the null hypothesis. Of course, the covariance matrices on which the reported test statistics are based derive from classical i.i.d. assumptions, but in Section 7 I consider alternative models.

Table 3 repeats the replication exercise for 8th grade reading scores. I view this outcome variable as more reliable for several reasons. First, the NELS sample offers considerably more observations in this wave. Second, despite "freshenings" of the sample in later waves, it seems likely that non-random attrition in follow-up surveys—both from dropouts and from student mobility—could bias the results. Finally, while the NELS students are random samples of their 8th grade schools, they are not representative of the schools they attend in later years. For these reasons, I use the 8th grade specifications as the basis for my investigations into robustness in the following sections. As before, my replication estimates of the choice effect are small and uniformly insignificant; specifications that use the 1942 choice index as an instrument estimate choice effects very near to zero. None of the three IV models rejects the OLS specification.

I have also estimated models similar to those shown here for the remainder of the NELS' fifteen test scores (reading, mathematics, history/geography, science, and a composite score in each of three grades). Classical test statistics reject zero for only four of the 60 estimated coefficients, all from models for the history/geography score and half with negative point estimates.

6. Sensitivity to sample specification

The estimates presented thus far have used the CCD MSA codes as the sole source of information about NELS schools' locations. The sample changes slightly when erroneous MSA codes are repaired, by hand for obvious errors or by recourse to the other sources of geocoding information in the CCD and NELS. Table 4 considers the sensitivity to these changes of the estimated choice effect on 8th grade reading scores. Estimates from the base replication sample are presented in Row 2. Row 1 presents estimates from the most restrictive possible sample, which excludes 115 schools for which the CCD reports a PMSA code as if it were an MSA code.[16] Row 3 adds three schools to the sample for which the CCD MSA code is obsolete. Row 4 uses the CCD school address to correct MSA codes for 19 schools. Finally, Row 5 uses the NELS zip code variables as an independent source of information, repairing the assignments of 3 schools located in zip codes contained within MSAs not clearly indicated by the CCD address. As the zip code variables are the only source of information about NELS private schools' locations, this sample forms the baseline for the analysis of pooled public and private samples in the next section. Choice effect estimates using the streams variable as the sole instrument are not particularly sensitive to these changes, but other estimates are notably more negative in the repaired samples than in the baseline sample.

7. Private enrollment and selection bias

As discussed earlier, estimates from samples of public school students can be biased relative to choice's effect on public school productivity if choice has an effect on the selection of students into private schools. Let $f_m = E\left[e_{isdm} \mid i \text{ attends public school; } m\right]$ be the average e_{isdm} of public school students in market *m*. If private school students are positively (negatively) selected in market *m*, $f_m < (>) \, 0$. What is relevant for the current analysis, however, is not the level of f_m but the extent to which it varies with choice ($\partial f_m / \partial c_m$).

In her Tables 5 and 6, Hoxby demonstrates a significant negative effect of choice on private enrollment rates. She interprets this as evidence that "choice among public schools is a substitute for choice of private schools" (p. 1233), and suggests that students with involved parents and high family incomes may be more likely to exercise either type of choice. If these characteristics are

[16] The Denver PMSA, e.g., is reported as 2080, rather than the "correct" CMSA+PMSA code of 342080.

positively correlated with e_{isdm}, choice would appear to pull high-e_{isdm} students out of the private sector and into public schools, so $\partial f_m / \partial c_m > 0$. When equation (1) is estimated on a sample of public school students, it is straightforward to show that $plim\left[\hat{\beta}\right] = \beta + \partial f_m / \partial c_m > \beta$ for any estimator $\hat{\beta}$ that is consistent in an unselected sample.

There are two obvious ways to correct Hoxby's specification for this selection bias. First, one can estimate (1) on a sample that is representative of total public and private enrollment in each MSA.[17] Under fairly strong assumptions—including that private schools are not systematically better or worse than public schools; that competition has similar effects on the productivity of public and private schools; and that any peer effects are linear and additive, so that stratification does not have an independent effect on average scores—this produces an unbiased estimate of the choice effect on average school productivity (Hsieh and Urquiola 2003). An alternative is a control function approach (Heckman 1979; Garen 1984; Card and Payne 2002), which attempts to absorb selection bias by controlling directly for the selection process as summarized by the MSA private enrollment rate. If the selection bias is in the hypothesized direction, either strategy should produce a smaller (more negative) estimate of the choice effect.

Table 5 explores the first strategy. Including private schools in the sample requires modifying Hoxby's specification somewhat, as district-level (SDDB) covariates are not available for these schools and they may only be matched to MSAs using the NELS zip code variables. The first six rows of the table explore the impact of excluding the SDDB covariates, which are not needed for consistent estimation of β as long as the MSA-level aggregates of these variables are included in X_m. Rows 1 and 3 report estimates of the base specification from the base replication sample and from

[17] Hoxby (1994a) adopts this strategy to study the effect of private school competition on public school performance.

the zip-code matched sample of public schools, respectively. Rows 2 and 5 repeat these estimates without district-level covariates, while rows 3 and 6 add back to the samples 6 public schools previously excluded for lack of SDDB data. Excluding the SDDB covariates produces slightly more negative choice effects in each IV specification, while adding the 6 schools missing SDDB data has essentially no effect. Finally, row 7 adds 174 private schools in the NELS sample. This produces a slightly more negative effect in the "streams" specification, but the other three estimates are less negative with private schools included.

One explanation for these uneven results might be that there is substantial sampling variability in the MSA private enrollment rate in the NELS sample, which includes less than one private school per MSA. For this reason, I turn to the control-function approach in Table 6. First on the basic replication sample (in Panel A) and then on the zip-code-matched public school sample (Panel B), I modify Hoxby's basic specification by including a control for an inverse Mills ratio derived from the MSA-level private enrollment rate, measured from Census data.[18] The Mills ratio coefficients are uniformly negative: Higher private enrollment rates seem to reduce public school average scores, as would be expected if private school students are positively selected, though only three of the eight estimates are significant. The inclusion of the private enrollment control lowers the choice effect slightly in all eight specifications. These effects, though again far from significant, are again consistent with the hypothesized selection bias in Hoxby's specification, and suggest that estimates of her specification may be upward-biased estimates of the effect of interdistrict choice on public school productivity.

[18] The inverse Mills ratio is $\lambda(\pi_j) \equiv \varphi\left(\Phi^{-1}(1-\pi_j)\right)\big/(1-\pi_j)$, where π_j is the MSA private enrollment rate (Card and Payne 2002). Specifications that instead include a polynomial in the private enrollment rate do not fit the data better than does the Mills ratio specification, and have similar consequences for the choice coefficient.

8. The clustered error structure and standard error estimates

I consider in this section several estimators of the sampling variance of the OLS and IV coefficients that are robust to the NELS data's clustered error structure. I am unable to reproduce Hoxby's standard errors. When I use her district and metropolitan area random effects specification, my estimated standard errors are 15-30% larger than hers; more robust error estimates range from 20-70% larger than those that she reports.

The NELS used a stratified sampling scheme, in which just over 1,000 schools were selected and 25 students were surveyed from each selected school. The 8th grade replication sample consists of 11,495 students, but they are drawn from only 548 schools, 432 districts, and 196 MSAs. If there are any unobserved factors influencing student test scores that are common to students in the same school, district, or metropolitan area, the classical standard errors considered so far—which impose $\mathrm{var}(e_{sdm}) = \mathrm{var}(e_{dm}) = \mathrm{var}(e_{m}) = 0$ in equation (3)—will likely overstate the effective sample size and understate sampling variances.

Table 7 reports several standard error estimates for the base replication sample and specification, as reported in Tables 2 and 3. Panel A considers the 8th grade reading score as an outcome, while Panel B presents results for 12th grade reading scores. Column A repeats the coefficients of interest from the earlier tables, along with their classical standard errors.

Columns B through D report estimates of the standard errors that derive from an implementation of Moulton's (1986) estimator for the variance of OLS and IV coefficients under an error components model like (2). In column B, only a metropolitan component is allowed (that is, I allow $\mathrm{var}(e_m) \neq 0$ but maintain $\mathrm{var}(e_{dm}) = \mathrm{var}(e_{sdm}) = 0$).[19] The standard deviation of e_m is 1.07 in

[19] This is a traditional random-effects model, although I estimate the sampling variance of classical OLS and IV coefficients rather than using a more efficient ML or FGLS estimator.

the OLS model in Panel A and similar (though not reported in Table 7) in IV models. Standard errors in Column B range from one-third to two-thirds larger than in Column A.

Column C loosens the error specification, allowing both metropolitan and district error components. This is the model that Hoxby uses. The estimated district component is positive, and drives out the MSA-level component (in Panel A, though not in Panel B). Standard errors of the choice coefficients are similar to those in Column B, but are substantially larger than those that Hoxby reports for the same model (repeated in Column G).

Column D reports an even less restrictive specification that allows for a school component to the error term. In both panels, the school component is quite large, and in its presence there are no estimated district or metropolitan area components.[20] Standard errors are slightly larger than in Column C.

Finally, Columns E and F explore less parametric variance estimators that are robust to heteroskedasticity and to arbitrary correlations among the residuals of students in the same metropolitan area. Column E reports the "clustered" standard errors that were also reported in Tables 2 and 3. These derive from a procedure that is consistent under arbitrary within-MSA covariance matrices, and therefore nests the error components model.[21] Clustered standard errors are similar to those in Columns C and D. A similarly non-parametric approach is to collapse the data to the metropolitan level, regressing average NELS scores in the MSA on the choice index, MSA-level controls, and the MSA average of all district, school, and individual controls. I present

[20] Note that there are relatively few districts in the sample that contribute multiple schools, providing little power for the district component estimation. Also, the 12th grade sample in Panel B is endogenously dispersed across schools, and the large school component may reflect this sorting as much as it does variation in school quality.

[21] This approach is asymptotically justified, with asymptotics in the number of clusters. It can be poorly behaved when the number of groups is small, but the 196 MSAs in the NELS sample are probably sufficient. Indeed, clustering is probably better behaved than is the Moulton estimator in situations, like that in Column D, where there are relatively few contrasts from which to estimate district-level error components.

estimates of this sort in Column F, with heteroskedasticity-robust standard errors. The estimated standard errors here are slightly larger than in previous columns.

Taken together, the estimates in Table 7 indicate that classical estimators substantially understate the sampling variability of the coefficient of interest in models like Hoxby's. All of the more robust estimators produce substantially larger standard errors than are estimated under the classical assumption of zero autocorrelation. This can have important implications for inference. Compare the standard errors in Table 8 with the point estimate that Hoxby reports for the effect of choice on 8th grade scores, 3.82. Only one of the five robust standard error estimates for the streams specification permits us to distinguish Hoxby's estimate from zero effect at the 5% level. This is not quite true for the 12th grade analysis, where Hoxby's coefficient of 5.77 has a *t*-statistic between 2 and 2.2 with four of the five robust standard errors. The replication estimate of this coefficient, of course, is substantially smaller than is Hoxby's, and has a *t*-statistic of about 1.

9. Evidence from SAT Scores

Although the NELS provides a random sample of students and schools, it is somewhat thin: There are fewer than 3 NELS public schools in the average MSA observed in the data, and over 100 MSAs contain no sample schools. The asymptotic distribution of the instrumental variables estimator may thus be a poor guide to its performance in the NELS sample. Table 8 presents estimates from an alternative data set containing observations on approximately one third of metropolitan SAT-takers from the 1994 cohort matched to their schools and MSAs. These data contain observations from nearly every school in the 190 MSAs in SAT-taking states. I use only observations from public schools, and modify Hoxby's specification by including a control for an inverse Mills ratio in the MSA private enrollment rate. Standard errors are clustered at the MSA level.

Although the SAT sample is thick, it is not random: Students choose whether to take the SAT, and the average SAT-taking rate in the sample MSAs is about 45%. In an effort to avoid sample selection bias coming from this decision, I also control for a quadratic in the MSA public school SAT-taking rate. For comparability to earlier results, SAT scores are standardized to the same mean (50) and standard deviation (10) as the NELS scores.

None of the four models in Table 8 indicates a statistically significant effect of choice on student SAT scores. Moreover, three of the four point estimates are negative. Exogeneity tests fail to reject the negative OLS coefficient in any of the IV specifications. Finally, the point estimates are quite small: Where Hoxby's 8th grade model indicates that a one unit increase in competition raises scores by 0.38 standard deviations, the effects in Table 9 range from -0.03 to +0.12, with an upper bound to the OLS confidence interval of +0.06.

10. Conclusion

The current analysis has not considered Hoxby's results from the NLSY, and cannot speak to the validity of her conclusions from those data, which echo her NELS results in indicating a salutary effect of interdistrict competition on school output. Hoxby seems to find her NELS estimates the most compelling, however, and focuses her discussion on these.

I have not been able to determine the source of the substantial divergence between Hoxby's (2000) results and my own. The research design requires bringing together over half a dozen distinct data sets, several of which exist in multiple versions or vintages and few of which are designed for analyses of this sort. The researcher must make many decisions, many arbitrary, about the creation of a sample and the definition of control variables; it would have been quite difficult for Hoxby to report these decisions in sufficient detail to permit exact replication. I have attempted to explore the sensitivity of the results in the analyses reported here and in many others that are not, but have not exhausted the set of possible permutations. Evidently, Hoxby made different decisions—surely equally defensible—than did I, with important consequences for the results.

The estimates from my reanalysis of the NELS data, along with the SAT score results, do not support Hoxby's conclusions that "naïve estimates (like OLS) that do not account for the endogeneity of school districts are biased toward finding no effects" (p. 1236). I find little evidence, in either data set, of bias in OLS estimates. OLS estimates are consistently more negative than are those from IV, but the difference is always well within the margin of sampling error.

There is similarly little evidence in the results presented here to support Hoxby's claim that "Tiebout choice raises productivity by simultaneously raising achievement and lowering spending" (p. 1236-7). I have not revisited Hoxby's results on the relationship between choice and school spending, but I am unable to find any discernable effect of choice on student achievement. A more conservative conclusion would be that the relatively few United States metropolitan areas, variation among which identifies the choice effects in the models considered here, are insufficient to precisely estimate any weak relationship that may exist between jurisdictional fragmentation and student performance. There is no basis here to reject a hypothesis of no effect, but neither is there much power against alternative claims of small effects, either positive or negative.

References

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.

Brennan, Geoffrey and James Buchanan (1980). The Power to Tax: Analytical Foundation of a Fiscal Constitution. Cambridge, Cambridge University Press.

Card, David and A. Abigail Payne (2002). "School Finance Reform, the Distribution of School Spending, and the Distribution of SAT Scores." Journal of Public Economics **83**(1): 49-82.

Chubb, John and Terry M. Moe (1990). Politics, Markets, and America's Schools. Washington, D.C., The Brookings Institution.

Davidson, Russell and James G. MacKinnon (1993). Estimation and Inference in Econometrics. New York, Oxford University Press.

Fiske, Edward B. and Helen F. Ladd (2000). When Schools Compete: A Cautionary Tale. Washington, D.C., Brookings Institution Press.

Friedman, Milton (1962). Capitalism and Freedom. Chicago, University of Chicago Press.

Garen, John (1984). "The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable." Econometrica **52**(5): 1199-1218.

Gray, E. R. (1944). Governmental Units in the United States 1942. Washington, D.C., Bureau of the Census.

Greene, William H. (2000). Econometric Analysis. New Jersey, Prentice Hall.

Heckman, James J. (1979). "Sample Selection Bias as a Specification Error." Econometrica **47**: 153-161.

Hoxby, Caroline M. (1994a). "Do Private Schools Provide Competition for Public Schools?" National Bureau of Economic Research: Working Paper #4978.

--- (1994b). "Does Competition among Public Schools Benefit Students and Taxpayers?" National Bureau of Economic Research: Working Paper 4979.

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

Hsieh, Chang-Tai and Miguel Urquiola (2003). "When Schools Compete, How Do They Compete? An Assessment of Chile's Nationwide School Voucher Program," Mimeo, Columbia University, August.

Krueger, Alan B. and Pei Zhu (2003). "Another Look at the New York City School Voucher Experiment." Princeton University: Education Research Section Working Paper #1.

Moulton, Brent R. (1986). "Random Group Effects and the Precision of Regression Estimates." Journal of Econometrics **32**(3): 385-397.

Rothstein, Jesse M. (2003). "Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions." Princeton University, Education Research Section: Working Paper #3.

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

Urquiola, Miguel (1999). "Demand Matters: School District Concentration, Composition, and Educational Expenditure." University of California, Berkeley, Center for Labor Economics: Working Paper #14.

Table 1. First-stage models for Hoxby's index of choice among districts

	Hoxby, Table 2	Replication			
	(A)	(B)	(C)	(D)	(E)
Instruments					
Number of larger streams / 100	0.080				
	(0.040)				
Number of smaller streams / 100	0.034		0.037		0.020
	(0.007)		(0.007)		(0.006)
1942 choice index				0.694	0.660
				(0.046)	(0.047)
Control variables					
Population (10,000,000s)	150.00	0.36	0.17	0.12	0.02
	(130.00)	(0.15)	(0.15)	(0.12)	(0.12)
Land Area (100,000s of sq. miles)	0.50	1.00	0.53	0.64	0.41
	(0.50)	(0.26)	(0.27)	(0.20)	(0.21)
Mean log(HH income)	-0.25	0.15	0.17	0.19	0.20
	(0.16)	(0.11)	(0.11)	(0.08)	(0.08)
Gini coeff., HH income	-3.58	-2.44	-2.44	-1.45	-1.48
	(0.81)	(0.80)	(0.77)	(0.60)	(0.59)
Population: Fr. aged 0-19	1.86	1.27	1.75	0.87	1.17
	(0.80)	(0.87)	(0.85)	(0.65)	(0.65)
Population: Fr. aged 65+	0.45	-0.21	0.37	0.74	1.02
	(0.61)	(0.72)	(0.70)	(0.55)	(0.55)
Population: Fr. Black	0.83	0.83	0.65	0.53	0.46
	(0.41)	(0.37)	(0.36)	(0.28)	(0.28)
Population: Fr. Hispanic	0.09	0.39	0.35	0.28	0.26
	(0.18)	(0.25)	(0.25)	(0.19)	(0.19)
Population: Fr. Asian	1.88	-0.77	-0.77	0.18	0.33
	(1.00)	(0.45)	(0.43)	(0.71)	(0.69)
Index of racial homogeneity	-0.18	0.71	0.40	0.40	0.27
	(0.49)	(0.29)	(0.29)	(0.23)	(0.23)
Index of ethnic homogeneity	0.70	-0.51	-0.40	-0.42	-0.36
	(0.70)	(0.30)	(0.29)	(0.22)	(0.22)
Adults: Fr. some college	-1.00	-1.85	-1.47	-0.79	-0.63
	(0.41)	(0.46)	(0.45)	(0.35)	(0.35)
Adults: Fr. BA+	0.95	0.31	0.37	0.03	0.06
	(0.42)	(0.35)	(0.34)	(0.27)	(0.26)
Index of educational homogeneity	-2.95	-1.47	-0.71	0.06	0.36
	(1.03)	(1.01)	(0.98)	(0.78)	(0.77)
N	316	335	335	333	333
R^2	nr	0.47	0.51	0.70	0.71
F-statistic for exclusion of instruments	24.4		24.8	223.7	121.9

Notes: Observations are unweighted metropolitan areas. Dependent variable in all columns is a Herfindahl-based index of choice among school districts in the MSA; see text for details. All models include 9 Census division fixed effects. "nr"=Not reported in Hoxby (2000). Hoxby's larger and smaller streams variables have mean (S.D.) 7.9 (14.8) and 182.7 (208.8), respectively. In the replication sample, these statistics are 193.4 (204.0) for smaller streams and 0.86 (0.26) for the 1942 choice index.

Table 2. Basic models for NELS 12th grade reading score, Hoxby (2000) and replication

	Hoxby (2000)		Replication			
	OLS	IV: Streams	OLS	IV: Sm. streams	IV: 1942 choice	IV: Both
	(A)	(B)	(C)	(D)	(E)	(F)
Index of choice among districts	-1.43	5.77	-0.16	2.93	0.96	1.13
	(0.65)	(2.21)	(0.55)	(1.67)	(0.78)	(0.77)
			[1.01]	[2.63]	[1.62]	[1.57]
Individual Covariates						
log(Family income)	nr	1.54	0.83	0.81	0.76	0.76
		(0.16)	(0.14)	(0.14)	(0.14)	(0.14)
Female	nr	1.96	1.95	1.91	1.98	1.97
		(0.23)	(0.23)	(0.23)	(0.23)	(0.23)
Asian	nr	0.28	-1.49	-1.47	-1.61	-1.62
		(0.59)	(0.60)	(0.60)	(0.62)	(0.62)
Black	nr	-5.49	-5.71	-5.57	-5.68	-5.67
		(0.50)	(0.43)	(0.44)	(0.43)	(0.43)
Hispanic	nr	-2.87	-2.33	-2.28	-2.40	-2.40
		(0.52)	(0.53)	(0.53)	(0.53)	(0.53)
Parents: Some college	nr	2.31	2.24	2.30	2.23	2.24
		(0.30)	(0.29)	(0.30)	(0.30)	(0.30)
Parents: BA +	nr	5.45	6.38	6.44	6.37	6.37
		(0.30)	(0.34)	(0.34)	(0.34)	(0.34)
District Covariates						
Mean log(HH income)	nr	nr	1.67	2.41	2.56	2.61
			(1.35)	(1.40)	(1.36)	(1.36)
MA-level Covariates						
Mean log(HH income)	nr	-5.42	-0.51	-1.16	-1.00	-1.05
		(5.53)	(1.70)	(1.74)	(1.72)	(1.72)
Population: Fr. Asian	nr	-5.62	-10.52	-9.64	-30.68	-30.78
		(13.07)	(10.49)	(10.53)	(11.68)	(11.68)
Population: Fr. Black	nr	-0.73	-0.90	-3.46	-3.34	-3.50
		(6.06)	(6.18)	(6.33)	(6.20)	(6.19)
Population: Fr. Hispanic	nr	0.25	4.40	2.59	2.65	2.56
		(3.52)	(4.92)	(5.02)	(4.94)	(4.94)
Adults: Fr. BA+	nr	3.16	3.64	3.05	5.48	5.47
		(5.93)	(3.92)	(3.94)	(3.94)	(3.94)
Number of students / MSAs	6,119 / 209		5,865 / 212			
Exog. test (*p*-value)	--	nr	--	0.05	0.14	0.06

Notes: Dependent variable is the NELS 12th grade reading score. Regressions use NELS weights, adjusted to weight each MSA equally. Following Hoxby, all columns include controls for MA-level population, land area, and 9 census division fixed effects, as well as for both district- and MA-level mean log household income and Gini coefficient; population fraction Asian, Hispanic, and Black; fraction of adults with some college and with BAs; and indices of racial, ethnic, and educational homogeneity. "nr"=Not reported in Hoxby (2000). Replication columns C-F report classical standard errors in parentheses, not the error-components estimates reported by Hoxby. Square brackets report "clustered" (on the MSA) standard errors for the choice coefficient. Using either errors, none of the unreported coefficients (except division effects) are significant in any of the replication specifications. Reported *p*-values for exogeneity tests are based on the classical standard errors.

Table 3. Basic models for NELS 8th grade reading score, Hoxby (2000) and replication

	Hoxby (2000)		Replication			
	OLS	IV: Streams	OLS	IV: Sm. streams	IV: 1942 choice	IV: Both
	(A)	(B)	(C)	(D)	(E)	(F)
Index of choice among districts	-0.24	3.82	-0.47	0.51	-0.02	0.03
	(0.49)	(1.59)	(0.42)	(1.36)	(0.61)	(0.60)
			[0.64]	[2.08]	[1.14]	[1.12]
Individual Covariates						
log(Family income)	nr	nr	1.01	1.00	1.00	1.00
			(0.09)	(0.09)	(0.09)	(0.09)
Female	nr	nr	2.11	2.11	2.09	2.09
			(0.17)	(0.17)	(0.17)	(0.17)
Asian	nr	nr	-0.58	-0.56	-0.60	-0.60
			(0.51)	(0.51)	(0.52)	(0.52)
Black	nr	nr	-4.09	-4.08	-4.09	-4.09
			(0.32)	(0.32)	(0.32)	(0.32)
Hispanic	nr	nr	-2.24	-2.23	-2.20	-2.20
			(0.37)	(0.37)	(0.37)	(0.37)
Parents: Some college	nr	nr	2.52	2.53	2.53	2.53
			(0.21)	(0.21)	(0.21)	(0.21)
Parents: BA +	nr	nr	6.92	6.94	6.94	6.94
			(0.25)	(0.25)	(0.25)	(0.25)
District Covariates						
Mean log(HH income)	nr	nr	4.29	4.52	4.37	4.39
			(0.99)	(1.03)	(1.00)	(1.00)
MA-level Covariates						
Mean log(HH income)	nr	nr	-3.36	-3.70	-3.51	-3.54
			(1.24)	(1.32)	(1.27)	(1.27)
Population: Fr. Asian	nr	nr	-14.63	-14.05	-13.58	-13.57
			(7.95)	(7.99)	(8.90)	(8.90)
Population: Fr. Black	nr	nr	-0.08	-0.68	-0.24	-0.28
			(4.31)	(4.39)	(4.36)	(4.36)
Population: Fr. Hispanic	nr	nr	2.34	1.89	2.24	2.22
			(3.46)	(3.52)	(3.51)	(3.51)
Adults: Fr. BA+	nr	nr	6.12	5.97	5.95	5.95
			(2.81)	(2.82)	(2.85)	(2.85)
Number of students / MSAs	10,790 / 211		11,495 / 196			
Exog. test (*p* -value)	--	nr	--	0.45	0.23	0.16

Notes: Dependent variable is the NELS 8th grade reading score. See notes to Table 2.

Table 4: Robustness of estimated choice effect on 8th grade reading scores to small changes in sample.

	Information used to match NELS schools to MSAs	Sample size:			Estimated Choice effect			
		MAs	Schools	Students	OLS	IV		
						Streams	Choice42	Both
		(A)	(B)	(C)	(D)	(E)	(F)	(G)
1.	Raw CCD MSA codes	163	433	9,098	-0.83	0.97	-0.38	-0.23
					(0.45)	(1.28)	(0.65)	(0.63)
					[0.66]	[1.99]	[1.16]	[1.15]
2.	PMSA codes on CCD repaired (*Base replication sample*)	196	548	11,495	-0.47	0.51	-0.02	0.03
					(0.42)	(1.36)	(0.61)	(0.60)
					[0.64]	[2.08]	[1.14]	[1.12]
3.	Obsolete/invalid CCD codes repaired	196	551	11,558	-0.48	0.47	-0.02	0.03
					(0.42)	(1.36)	(0.61)	(0.60)
					[0.64]	[2.09]	[1.14]	[1.12]
4.	District address used to assign MSA	195	553	11,614	-0.78	0.51	-0.53	-0.43
					(0.43)	(1.35)	(0.61)	(0.60)
					[0.67]	[1.99]	[1.16]	[1.14]
5.	Zip code used to confirm district address	195	556	11,683	-0.77	0.58	-0.60	-0.49
					(0.42)	(1.35)	(0.61)	(0.60)
					[0.67]	[1.99]	[1.15]	[1.13]

Notes: Each row explores a different method of assigning NELS schools to metropolitan areas. 115 schools are assigned differently in row 2 than in row 1; 3 in 3 than in 2; 19 in 4 than in 3; and 3 in 5 than in 4. Specifications are those in Table 3. Classical standard errors in parentheses; clustered standard errors in square brackets.

Table 5: Estimated choice effect when sample includes private schools

Specification	Sample Size	Estimated Choice effect			
		OLS	IV		
			Streams	Choice42	Both
	(A)	(B)	(C)	(D)	(E)
Panel A: Basic replication sample					
1. Hoxby specification	11,495	-0.57	0.17	-0.15	-0.11
		(0.41)	(1.31)	(0.60)	(0.59)
		[0.63]	[1.99]	[1.12]	[1.09]
2. Without district-level (SDDB) controls	11,495	-0.74	-0.18	-0.41	-0.38
		(0.41)	(1.27)	(0.60)	(0.58)
		[0.63]	[1.88]	[1.09]	[1.06]
3. Add 6 public schools without SDDB data	11,627	-0.71	-0.18	-0.36	-0.34
		(0.41)	(1.27)	(0.60)	(0.58)
		[0.63]	[1.91]	[1.09]	[1.06]
Panel B: Zip code-matched sample					
4. Hoxby specification	11,683	-0.81	0.27	-0.68	-0.58
		(0.42)	(1.31)	(0.61)	(0.59)
		[0.66]	[1.92]	[1.14]	[1.11]
5. Without district-level (SDDB) controls	11,683	-0.88	-0.08	-0.87	-0.79
		(0.41)	(1.27)	(0.60)	(0.59)
		[0.66]	[1.88]	[1.11]	[1.08]
6. Add 6 public schools without SDDB data	11,815	-0.83	-0.07	-0.78	-0.71
		(0.41)	(1.28)	(0.60)	(0.59)
		[0.66]	[1.90]	[1.10]	[1.07]
7. Add 174 private schools	15,298	-0.25	-0.17	-0.34	-0.32
		(0.36)	(1.10)	(0.51)	(0.50)
		[0.64]	[1.82]	[1.01]	[1.00]

Notes: Sample and specification in Rows 1 and 4 is that from Table 4, rows 1 and 5, respectively. Rows 2 and 5 exclude the district-level controls from Hoxby's specification. Rows 3 and 6 add to the sample 6 schools previously excluded for lack of district-level covariates. Row 7 adds an additional 174 private schools that are excluded from all previous estimates. Classical standard errors in parentheses; clustered standard errors in square brackets.

Table 6: Estimated choice effect when private enrollment is controlled

	Estimated choice effect				Inverse Mills ratio coefficient			
	OLS	IV			OLS	IV		
		Streams	Choice42	Both		Streams	Choice42	Both
	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
Panel A: Basic replication sample								
1. Hoxby specification	-0.47	0.51	-0.02	0.03				
	(0.42)	(1.36)	(0.61)	(0.60)				
	[0.64]	[2.08]	[1.14]	[1.12]				
2. With control for private enrollment	-0.56	0.35	-0.14	-0.09	-2.70	-2.16	-2.46	-2.43
	(0.42)	(1.41)	(0.62)	(0.61)	(1.82)	(1.99)	(1.85)	(1.85)
	[0.62]	[2.11]	[1.13]	[1.10]				
Panel B: Zip code-matched sample								
3. Hoxby specification	-0.77	0.58	-0.60	-0.49				
	(0.42)	(1.35)	(0.61)	(0.60)				
	[0.67]	[1.99]	[1.15]	[1.13]				
4. With control for private enrollment	-0.90	0.36	-0.76	-0.66	-3.96	-3.20	-3.91	-3.84
	(0.43)	(1.40)	(0.62)	(0.61)	(1.83)	(2.00)	(1.85)	(1.85)
	[0.65]	[1.99]	[1.14]	[1.12]				

Notes: Rows 1 and 3 report estimates of choice effect in Hoxby's specification from Table 4, rows 1 and 5, respectively. Rows 2 and 3 add a control for an inverse Mills ratio derived from the MSA-level private enrollment rate to the Hoxby specification, using the same samples as in the preceeding rows. Classical standard errors in parentheses; clustered standard errors in square brackets.

Table 7. Alternative standard error estimates for 8th grade replication sample

	Replication estimates and standard errors						Hoxby (2000)
	Classical	Error components			Cluster on MA	Data collapsed to MA level	
		MA	MA & District	MA, Dist. & School			
	(A)	(B)	(C)	(D)	(E)	(F)	(G)
Panel A: Base year (8th grade) replication sample							
Coefficients and estimated standard errors							
OLS	-0.47					-0.08	-0.24
	(0.42)	(0.56)	(0.56)	(0.59)	(0.64)	(0.77)	(0.49)
IV: Streams	0.51					0.98	3.82
	(1.36)	(1.96)	(1.85)	(2.13)	(2.08)	(2.57)	(1.59)
IV: Choice42	-0.02					0.15	
	(0.61)	(1.02)	(1.09)	(1.22)	(1.14)	(1.29)	
IV: Both	0.03					0.22	
	(0.60)	(0.98)	(1.02)	(1.16)	(1.12)	(1.26)	
Standard deviation of error components (OLS model)							
MA (N=196)		1.07	0.00	0.00			nr
District (N=432)			1.37	0.00			nr
School (N=548)				1.96			
Individual	8.95	8.83	8.82	8.83			
Panel B: Second follow-up (12th grade) replication sample							
Coefficients and estimated standard errors							
OLS	-0.16					0.19	-1.43
	(0.55)	(0.72)	(0.79)	(0.86)	(1.01)	(1.11)	(0.65)
IV: Streams	2.93					3.08	5.77
	(1.67)	(2.73)	(2.88)	(3.35)	(2.63)	(2.74)	(2.21)
IV: Choice42	0.96					1.64	
	(0.78)	(1.27)	(1.48)	(1.76)	(1.62)	(1.86)	
IV: Both	1.13					1.76	
	(0.77)	(1.23)	(1.43)	(1.70)	(1.57)	(1.80)	
Standard deviation of error components (OLS model)							
MA (N=212)		1.65	1.27	0.00			
District (N=513)			1.90	0.00			
School (N=744)				3.64			
Individual	8.69	8.26	8.24	8.19			

Notes: Columns A-E report alternative estimates of the standard error of the choice coefficient from the replication estimates in Table 3, deriving from different assumptions about the correlation between the residual test scores of students in the same school, district, and MA. Column A assumes that all observations are independent, and reports classical standard errors. Columns B-D allow for components of the error at the MA; MA and district; and MA, district, and school levels, and estimate standard errors using Moulton's (1986) formula for random-effects-style models. Column E reports standard error estimates from STATA's "cluster" command, which is robust to arbitrary heteroskedasticity and within-MA serial correlation. Column F re-estimates the Hoxby model on MSA-level means, implicitly permitting arbitrary within-MA serial correlation, with heteroskedasticity-robust standard errors. Column G reports Hoxby's estimates, which she reports allow for district and MA error components as in Column C.

Table 8. Estimates from SAT data

	OLS	IV: Sm. streams	IV: 1942 choice	IV: Both
	(A)	**(B)**	**(C)**	**(D)**
Index of choice among districts	-0.26	1.23	-0.29	-0.17
	(0.43)	(1.25)	(0.56)	(0.54)
Individual Covariates				
log(Family income)	1.37	1.37	1.36	1.36
	(0.07)	(0.07)	(0.07)	(0.07)
Female	-1.26	-1.26	-1.27	-1.27
	(0.09)	(0.09)	(0.09)	(0.09)
Asian	0.59	0.56	0.59	0.59
	(0.25)	(0.25)	(0.26)	(0.26)
Black	-6.61	-6.62	-6.62	-6.62
	(0.16)	(0.16)	(0.16)	(0.16)
Hispanic	-2.99	-3.00	-2.99	-2.99
	(0.14)	(0.14)	(0.14)	(0.14)
Parents: Some college	1.79	1.79	1.79	1.79
	(0.12)	(0.12)	(0.12)	(0.12)
Parents: BA +	4.79	4.79	4.79	4.79
	(0.14)	(0.14)	(0.14)	(0.14)
District Covariates				
Population: Fr. Asian	-5.47	-4.17	-4.78	-4.77
	(2.67)	(2.86)	(3.03)	(3.04)
Population: Fr. Black	-4.02	-3.86	-3.88	-3.89
	(1.54)	(1.59)	(1.58)	(1.58)
Population: Fr. Hispanic	-4.12	-3.98	-4.00	-4.01
	(1.59)	(1.67)	(1.59)	(1.60)
Adults: Fr. some college	4.90	5.01	4.99	5.00
	(1.80)	(1.84)	(1.82)	(1.82)
Adults: Fr. BA+	8.73	8.26	8.65	8.62
	(1.35)	(1.48)	(1.38)	(1.38)
Index of educational homogeneity	7.54	7.38	7.43	7.43
	(2.16)	(2.13)	(2.16)	(2.15)
Exogeneity test (*p*-value)		0.15	0.78	0.57

Notes: Dependent variable in all columns is the SAT score, re-scaled to have mean (S.D.) 50 (10). Sample consists of SAT takers attending public schools in metropolitan areas wholly contained within "SAT states." N=332,029 students; 4,453 schools; 2,546 districts; and 190 MSAs. Regressions use SAT sampling weights, adjusted to weight each MSA equally. Standard errors, clustered on the MSA, in parentheses. Specification includes all control variables listed in notes to Table 2, as well as an inverse Mills ratio derived from the MSA-level private enrollment rate and a quadratic in the MSA public school SAT-taking rate. None of the suppressed control variable coefficients is significant in any specification, except land area in Column B and several division effects.