Appendices to:

"Does Competition among Public Schools Benefit Students and Taxpayers? A Comment on Hoxby (2000)"

Jesse Rothstein

December 2004

Appendix A: Data

1. The Hoxby/NCES data

Each issue of the *American Economic Review* includes a paragraph describing the "Policy on Data Availability:" "…to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication." Despite repeated requests over several years, Hoxby has not provided the data used to generate the results in her published paper (Hoxby, 2000). She has, however, provided a corrected data set (Hoxby, 2004a) to the National Center of Educational Statistics, which will distribute it to researchers with licenses for the confidential version of the National Educational Longitudinal Survey. I refer to these data throughout as the "Hoxby/NCES data." Hoxby (2004c) notes several differences between the data used in her paper and the corrected Hoxby/NCES data. First, where she originally used MSA codes from the School District Data Book (SDDB) to assign school districts to metropolitan areas, she now uses the Common Core of Data (CCD), with a few manual corrections, for this purpose. Second, "[t]he published version of the paper showed a version of the first stage regression that was outdated. The paper went through multiple revisions and Table 2 evidently was not updated so that it corresponded exactly with Tables 3+." As I document in the main text, the actual first stage regression is substantially different than that indicated in the published paper.

Hoxby constructs her new district-MSA crosswalk from the Common Core of Data agency files, using a field on these files that indicates the MSA in which each district is located. Unfortunately, there are substantial errors in this field. Many of these come from a small number of MSAs whose boundaries have changed over time; the CCD often contains obsolete MSA codes for districts in these areas.[1] One example is the "Kansas City, MO-KS MSA," which was coded 3760 in 1990 but had been divided into separate Missouri (code 3760) and Kansas (code 3755) components in 1983. Even the 1991-92 CCD, the most recent used for Hoxby's crosswalk, reports the obsolete 3755 code for 19 districts. Hoxby corrects six of these in Johnson County, Kansas, but retains the incorrect code for 13 districts in other counties.[2] Another common occurrence is a metropolitan district for which the CCD is simply missing an MSA code. One such district is the Collier County School District, which serves the entire Naples, Florida MSA (5345); as a result of this miscoding, the Naples MSA does not show up in Hoxby's analyses. A final sort of error involves MSA codes that are flatly incorrect: The CCD contains MSA codes of 6640, corresponding to the Raleigh-Durham MSA, for several non-metropolitan school districts in Ohio. As a result, four of the 11 districts in the Hoxby/NCES district-level data set that appear to be in Raleigh are actually in Ohio.

In addition to her data, Hoxby provided the Stata program that she wrote to assemble it. This program, entitled "construct.do," is included on the Hoxby/NCES CD. In examining this program, I discovered several glitches. A few of the most important are enumerated here:

1. Hoxby merges the NELS student-level file to the NELS school-level file, first merging on the base year school ID (*sch_id*), next the first follow-up school ID (*f1sch_id*), and finally the second follow-up ID (*f2sch_id*). The relevant section of her code (lines 25-66[3]) is reproduced in Table A1. Note that Hoxby does not rename any variables from the school-level data set between merges, nor does she specify the "update" option on her second and third merges (lines 52 and 62). As a result, nothing is changed by these latter two merges: All variables from the "using" data set (the school-level data) already exist on the "master" (student-level) data, and Stata does not alter pre-existing variables on the master data without explicit instruction.[4]

---

[1] Hoxby writes that "the Common Core files use the 1990 MSA and PMSA codes, even though some files have names that include years that predate 1990. This conclusion is based on inspection of the counties that were not included in the 1980 MSAs and were included in the 1990 MSAs" (Hoxby, 2004b, lines 506-8). In fact, the documentation to the 1987-88 file, from which Hoxby takes many of her MSA codes, includes several pages titled "Alphabetical Listing of Metropolitan Statistical Areas, October 1984…" (http://www.nces.ed.gov/ccd/pdf/pau87gen.pdf), suggesting that the compilers of that file did not intend to be using up-to-date 1990 codes. Inspection of the CCD files indicates that the MSA codes used do not derive from any single set of MSA definitions, but are drawn from several different generations of those definitions.

[2] Hoxby also reports a "larger streams" count for the 3755 MSA, though the erroneous CCD crosswalk should not have been necessary for the construction of this variable.

[3] All line numbers refer to lines in the "construct.do" program distributed with the September 2, 2004 generation of the Hoxby data.

[4] "[U]`pdate` varies the action `merge` takes when an observation is matched. By default, the master dataset is held inviolate—values from the master dataset are retained when variables are found in both datasets. If

2. Another problem appears in the same segment of code. Students and schools that entered the NELS sample in one of the follow-up surveys (11.5% of student observations and 57% of schools) have *sch_id* set to missing. Observations with missing *sch_id* are not excluded from either the master or the using data sets before the first merge is performed on line 42. Stata's merge command does not make an exception for missing values, so student and school observations with missing *sch_id* are assumed to correspond. Because there are many such observations on either side of the merge, Stata handles them in order, assigning the first such student to the first such school, the second to the second, and so on.[5] Thus, the resulting data set is incorrect—students with missing *sch_id* should not be assigned to any school on this merge—and its precise form depends on the order in which observations with missing *sch_id* appear in the two data sets. As it happens, each data set arrives at this point having been sorted on *sch_id* (lines 29 and 34), without explicit indication of how to sort observations within *sch_id* groups. Stata breaks ties randomly in sorts, so the order of the relevant observations is random, and indeed is different each time the program is run.[6]
3. A third, related glitch only affects the MSA-level data set, which Hoxby uses to estimate the MSA-level version of the first-stage. Hoxby's student-level models include dummy variables for the nine Census divisions, assigned on the basis of the school's location. The MSA-level analogue of this would use division variables that were not dummies, but which represented the fraction of enrollment in the MSA in each division (for those MSAs which span two or more divisions). Hoxby's computer program does not do this. Rather, it assigns a single division to each MSA, hand-coded for MSAs spanning divisions using the division in which the plurality of the MSA's enrollment is located. There are some MSAs, however, which are wholly contained within a single division but which are incorrectly assigned districts in other divisions by the erroneous CCD district-MSA crosswalk. For example, as noted earlier, four Ohio districts are assigned to the Raleigh-Durham metropolitan area. For MSAs that do not span divisions, division codes are assigned based on the first district in each MSA (line 2370), using a district data set sorted by the MSA code (*fipsmsa*, line 2369). Again, Stata breaks ties randomly in sorts, so which district happens to come up first within the Raleigh MSA is different each time the program is run, and on 36% (4/11) of iterations Raleigh is assigned to the East North Central division (containing Ohio).[7]

`update` is specified, however, the values from the using dataset are retained in cases where the master dataset contains missing," (StataCorp, 2003, "Reference G-M," p. 435). Hoxby apparently also intended `update`-style merges in lines 527-543.

[5] "When one is performing a match merge, the master and/or using datasets may have multiple observations with the same `varlist` value. These multiple observations are joined sequentially, as in a one-to-one merge. If the datasets have an unequal number of observations with the same `varlist` value, the last such observation in the *shorter* dataset is replicated until the number of observations is equal," (StataCorp, 2003, "Reference G-M," p. 446). Had Hoxby specified the `unique` or `uniqusing` options on her line 46, Stata would have refused to perform the match.

[6] "[S]`table`" specifies that observations with the same values of the variables in `varlist` are to keep their same relative order in the sorted data as they had previously….Without the `stable` option, the ordering of observations with equal values of `varlist` is randomized" (StataCorp, 2003, "Reference S-Z," p. 88).

[7] The Hoxby/NCES dataset happens to be one in which Raleigh is assigned to the wrong division.

4. MSA demographic characteristics are computed by summing observations on each district (from the SDDB) within the MSA. This creates two problems. First, any errors or omissions in the district-MSA crosswalk are reflected in measured MSA characteristics. Second, many people are double-counted: Any person living in an area served by separate elementary and secondary school districts will be counted twice toward the MSA totals.

To gauge the extent of the randomness introduced by errors 2 and 3, I executed Hoxby's program 10,000 times without alteration, tabulating the estimated choice effects (on the 12th grade reading sample) from each iteration. The distribution of estimates is displayed in Figure A1. The distribution is reasonably concentrated around its median (5.41), with a standard deviation of 0.47. However, the range is quite broad: The smallest estimated effect is 2.18, and the largest is 8.15.

2. Replication data sets

Given the above concerns, a complete replication required re-creating the analysis data set. My first step was to create a correct district-MSA crosswalk. Outside of New England, MSAs consist of whole counties. I used the CCD's county codes, which appear to be more reliable than its MSA codes, to assign districts to MSAs. In New England, town boundaries define MSAs and a district's county location is not sufficient to assign it to an MSA. I built the New England portion of the crosswalk by hand, examining each district's name and mailing address and using these to assign each to a town (and therefore an MSA). I have made available the code needed to produce this crosswalk from the public-use CCD data. While I cannot guarantee that the New England portion is completely free of errors and misjudgments, I believe that it is generally accurate.

I created two replication samples, using my repaired district-MSA crosswalk for each. My first replication sample follows Hoxby's algorithm (as expressed in her code) as closely as possible, but for the substitution of the improved crosswalk and the repair of the errors noted above. Repairing #4 required an alternate source of metropolitan demographic characteristics. I used the Summary Tape File 3A from the 1990 census[8] for this purpose, with one exception that is noted below.

Repairing #1 was also not straightforward. Each school in the NELS school file has up to six codes identifying its district (an "NCES" code and a "QED" code in each of three survey waves). Moreover, each student is assigned to up to three schools, one in each wave.

[8] This is based on the same raw 1990 long form census data as the SDDB, but avoids double-counting.

As a result, there are potentially 18 distinct district assignments for each student. By virtue of glitch #1, however, Hoxby observes only the six first-wave codes for each student. I used the following algorithm to assign students to schools:

1. For each school in each wave, use the NCES code in place of the QED code if both are valid district codes; otherwise, use the valid one. (Lines 369-374 of "construct.do" advocate resolving discrepancies between NCES and QED codes in favor of the NCES codes.) This narrows the 18 codes to 9.
2. Match each student to his/her school in each wave, and keep only that wave's district code from that school. This leaves a maximum of 3 district codes for each student: The BY code from the BY school, the F1 code from the F1 school, and the F2 code from the F2 school.
3. Follow the "majority rule" algorithm implemented in lines 416-491 of "construct.do" to resolve discrepancies among these three district codes, preferring districts appearing in two waves to one and districts appearing earlier to later. The "construct.do" algorithm neglects a few possibilities—e.g. students who attend different districts in two of the waves and are missing a district assignment in the third wave—but I applied the same rule for these.
4. Following lines 432-446 of "construct.do," assign students who are as yet unassigned to the modal district among successfully-assigned students who attend the same school as them in the BY, F1, or F2 waves. This has the effect of assigning some students who attended private schools in all waves to the public school that some of their private school classmates in, say, the BY wave transferred to in F1. (I omitted private school students from all analyses of this replication sample, however. As a result, this step is not particularly consequential.)

In constructing my second replication, I deviated from Hoxby's algorithm when what seemed to me more sensible options were available. I refer to this sample as the "preferred replication" sample. The most important difference is in the way that students are assigned to districts. In an effort to prioritize accuracy over maximizing the sample size, I used only contemporaneous information for this purpose, allowing individual respondents' district assignments to vary between waves. Thus, a student whose F2 school is missing a district code in the F2 wave was omitted from my analysis sample for F2 scores. (Hoxby's algorithm would use information about the 8th grade school if it were available, and indeed for all students would use the same district assignment in each wave.) In practice, this meant following steps 1 and 2 of the above algorithm, then constructing distinct student-district matches for each survey wave using only the district code for the school attended in that wave. There are also important differences in the construction of individual variables, as detailed below.

In Section IV, I also extend the preferred replication sample to include private school students. An alternative algorithm is required for these students, as district codes are unavailable for NELS private schools. I make use of information about the demographic characteristics of the zip codes in which schools are located. These variables are drawn from the zip code tabulation of the 1990 Census, so these data (the "STF-3B" file) can be used to assign zip codes to each NELS school. The vast majority of zip codes are contained entirely within a single county; for those that are not, I assign the zip code to the county containing the plurality of its population. The resulting crosswalk from NELS schools to counties is sufficient to assign schools to MSAs outside of New England. In New England, however, counties do not map to unique MSAs. Using mapping software, cartographic boundary files from the 1990 Census, and a third-party vendor's zip code boundary files (ESRI, 2002), I compute the fraction of each zip code's land area contained within each MSA, and assign each zip code to the MSA containing the plurality of its area. (Once again, for the vast majority of zip codes there are no ambiguities.) There are a very small number of zip codes containing NELS schools that cannot be assigned in this way, which I code by hand.

Note that this algorithm assigns each NELS school, public and private, to an MSA, but not to a district. For analyses of these data, I exclude the district-level covariates from Hoxby's specification. For consistency, I use the zip-code based assignments for both public and private schools, even though alternative district-based assignments (which agree in every case where both are defined) are available for the public schools.

All code for the assembly of my replication samples and for estimation of the models presented has been made available in my "makedata_msadist.do" and "makedata_nels.do" program files. Unfortunately, I cannot make the individual-level data available in the same manner, as they derive from a restricted-access version of the NELS data and are available only from NCES and only to licensed researchers. However, researchers with the appropriate licenses should be able to run my computer programs to extract my samples from the restricted-access NELS data. There are a very few lines of code which have been redacted from my computer programs, to avoid compromising the confidentiality of the NELS data. I have asked NCES to distribute this code—in a file, "confidential.include.zipcode.do," that is called from the programs that I do distribute—to licensed researchers who request it. In addition, I describe here the covariates included in Hoxby's specification and how they are constructed in each replication sample.

3. Comparisons of individual covariates

a. Covariate construction, individual level

I begin with the construction of individual-level variables in the NELS data, as these are the most straightforward. For each item, I describe any discrepancies between the Hoxby/NCES construction and my preferred construction. In each case, I followed the Hoxby/NCES algorithm for the "close replication" sample (variable names are prefixed "*ch_*") and my preferred algorithm for the "preferred replication" sample (prefix "*jr_*").

- Indicators for attending a public school in each of the three waves (*pubschby*, *pubschf1*, *pubschf2*): Unless otherwise noted, all analyses include only students enrolled in public school in the wave from which the test score data are taken. There are several NELS variables describing the school sector in each wave. In 40 cases, these are not mutually consistent for the base year. Hoxby does not resolve these inconsistencies; I drop these observations from both replication samples for analyses of base year scores.
- The log of family income (*ch_lnfaminc* and *jr_lnfaminc*): The NELS data report family income in bins. Hoxby assigns each family the log of the midpoint of the relevant bin (in thousands). She assigns a "midpoint" of $800 for the $1-$1,000 bin and one of $220,000 for the $200,000+ bin. Observations for which the family income is reported as zero are set to missing. *ch_lnfaminc* follows this algorithm. I use a slightly different construction for *jr_lnfaminc*: I assign each family the log of the geometric average of the endpoints of their bin. (That is, a family in the $1,000-$3,000 bin is assigned ln(2) by Hoxby's algorithm, and (ln(1000)+ln(3000))/2 in mine.) I assign ln(500) to families in the $1-$1,000 bin, ln(250,000) for families with incomes above $200,000, and ln(1) to families with zero income. Finally, I use the 2nd follow-up survey's income variable (*f2faminc*) to assign values for students with missing incomes in the base survey variable (*byfaminc*); Hoxby uses only the latter variable.

- Student race (*ch_asian*, *ch_hispanic*, *ch_black*, *jr_asian*, *jr_hispanic*, and *jr_black*): Hoxby uses only the base-year *race* variable to assign these, and sets each to zero for students with *race*=8 (race missing or more than one race reported). Each dummy is set to missing if there is no value for the *race* variable. I supplement this variable with analogues from the follow-up surveys (*f1race*, *f2race1*, *f2rrace1*) when it does not resolve the student's race, and I set the indicators to missing if none of the four survey variables indicate a specific race.
- Student gender (*ch_female*, *jr_female*): Again, Hoxby uses only the base wave *sex* variable, and sets *female* to missing if this variable is missing; I supplement the base wave with analogous variables from the follow-ups (*f1sex*, *f2sex*, *f2rsex*).[9]
- Parental education (*ch_parscol*, *ch_parcolg*, *jr_parscol*, *jr_parcolg*): Hoxby uses the *bys34a* and *bys34b* variables to assign each parent's education, then uses the highest of these to assign her variable. *bys34a* and *bys34b* are student reports of their parents' education in the base year survey. I use instead *bypared*, *f1pared*, and *f2pared*. These use parent reports where they are available, and student reports only when they are not. There are many discrepancies between these variables and the student reports; it seems likely that the parent reports are more accurate.
- Test scores: Hoxby's preferred specification takes the 12th grade reading score, *f22xrstd*, as the dependent variable, though she also reports results for 8th grade reading (*by2xrstd*) and 10th grade mathematics (*f12xmstd*). Each score is normalized to have a mean around 50 and standard deviation around 10. In the main text, I analyze *f22xrstd* and *by2xrstd*, the 12th and 8th grade reading scores. I also present estimates in this appendix for four additional scores: Mathematics in all three waves, and reading in 10th grade (*by2xmstd*, *f12xmstd*, *f22xmstd*, and *f12xrstd*).

Summary statistics for each of these variables are reported in Table A2, for Hoxby's data set and for each of the two replication samples. The rightmost columns of the Table report correlation coefficients between the different samples, computed pairwise over observations that have values in each of two samples. Note that most of these correlations are almost

[9] One might not expect many missing values for such a basic demographic characteristic. In fact, it is missing for over 10% of observations. These are students who were brought into the NELS sample via "freshenings" in the first and second follow-up surveys. By using only base-year variables, Hoxby excludes freshened observations. Her 12th grade sample is thus not representative of 12th graders in 1992, but only of 1988 8th graders who were in 12th grade in 1992.

exactly one (with the notable exception of the parental education variables). The primary differences between samples are in the number of observations with missing values—note, for example, that the standard deviation of ln(family income) is higher in the preferred replication sample than in either of the other samples, where it is defined for 3,000 fewer observations.

b. Covariate construction, district level

District covariates are taken from the School District Data Book (SDDB), a tabulation of 1990 Decennial Census data along school district boundaries. There are several extant versions of the SDDB, not all of which are complete.[10] The Hoxby/NCES CD includes a copy of the "Top 100" file (containing most of the variables that are used). There are some necessary variables which are not included in the "Top 100" file. The Hoxby/NCES CD provides extracts of these variables in two separate files. It is not clear what version of the SDDB was used for these, nor does is the code provided that was used to perform the extraction. The supplementary files have somewhat fewer records than does the (presumably complete) "Top 100" file.

In constructing my replication samples, I rely on a version of the SDDB data obtained from the contractor who produced the file for NCES.[11] The full data set is too large (1.4 GB zipped) to easily distribute, but I have made available my extraction program and can work with interested readers to help them obtain the data. There do not appear to be major differences between the two versions of the data.[12]

- Racial composition of the district (*ch_d_pop_fas*, *ch_d_pop_fbl*, and *ch_d_pop_fhi*; *jr_d_pop_fas*, *jr_d_pop_fbl*, and *jr_d_pop_fhi*): Hoxby excludes people of "other" race in her computations (so the denominator is the sum of the Hispanic and non-Hispanic black, white, Asian, and American Indian populations). I include the "other race" group in the denominator.

---

[10] The NBER, for example, has a version acquired from the National Archives at http://www.nber.org/sddb/. Documentation on that page reads "Currently we have and have online 156 files, which NARA says is the whole file, or least all they have. They believe there may be records, or parts of records missing from California and Minnesota. We have observed that Minnesota contains 603 undecipherable records."
[11] I am grateful to Cecilia Rouse and Lisa Barrow for providing me with these data.
[12] There are 357 districts that appear in Hoxby's SDDB extract but not in mine. Hoxby codes only two of these as metropolitan (one apparently incorrectly). Only 11 of the districts in question (and neither of the apparently metropolitan districts) have non-missing enrollment.

- Index of racial homogeneity (*ch_d_herfrace* and *jr_d_herfrace*): This is computed as the sum of each race's squared population share. As noted above, Hoxby's population shares exclude "other" race from the denominator, where mine include them. I also include "other" as one of the races over which the sum is taken.
- Educational distribution (*d_ed_scol*, *d_ed_ba*, and *d_herfed*): We use identical algorithms. *d_herfed* is the sum of the squared population shares of the less than high school, high school graduate, some college, and BA+ groups.
- Household income (*d_lnmeanincA*): Hoxby (2000) describes this variable as the "mean of log(household income) in the district." In fact, she constructs it as the log of the mean household income in the district. I follow this definition.[13]
- Gini coefficient of household income (*ch_d_gini* and *jr_d_gini*): These are constructed from the distribution of district households across 25 income bins. All households in each bin are assumed to have income equal to the midpoint of the bin. Hoxby assigns the bottom bin ($0-$5,000) a "midpoint" of $4,000 and the top bin ($150,000+) a value of $175,000. I use $2500 for the bottom bin. For the top bin, I use a variable describing the total income among families with incomes above $150,000 (*P81_2*) to construct the actual mean income among families in this bin in the district. Not surprisingly, this has important consequences for the Gini coefficient.
- Index of ethnic homogeneity (*ch_d_herfethn*, *jr_d_herfethn*): This is a modified version of the index of racial homogeneity, mentioned above. Hoxby (2000) does not provide the formula, but cites another paper (Alesina et al., 1999) for its construction. That paper describes homogeneity indices for the Hispanic and white populations, but does not describe how these are to be aggregated into a single ethnic homogeneity index. The formula, as used in the Hoxby/NCES code, is:

---

[13] Given the identical construction, one would expect this variable to be perfectly correlated between the Hoxby/NCES data set and the replication data sets (the fact that Hoxby divides average income by $1,000 before logging notwithstanding). While the correlation is quite high, it is not exactly one. There are two sources of income information on the SDDB: Variables tabulating the number of households in each of 25 income bins (named *P80*), and variables recording the aggregate income among families with incomes above and below 150,000 (named *P81*). I use the latter variables to construct mean household income, but when I use the former (and Hoxby's rules, used elsewhere in her code, for assigning households to midpoints of each bin) I replicate her variable to within four decimal places for all but 39 districts. Assuming that the census tabulates each variable correctly, the "aggregate income" approach is almost certainly more accurate than the "bin midpoints" approach.

$$\text{Index}=(\text{FrBl})^2 + (\text{FrAs})^2 + (\text{FrIn})^2 + (\text{FrWh})^{3\text{-WhIndex}} + (\text{FrHi})^{3\text{-HiIndex}} + (\text{FrOt})^2,$$

where FrBl, FrAs, FrIn, FrWh, FrHi, and FrOt are the population shares black, Asian, American Indian, white, Hispanic, and other race. (As noted above, Hoxby omits the final group). WhIndex and HiIndex are indices of the source-country heterogeneity of the white and Hispanic populations, respectively. The white index is the sum of the squared shares of the white population that are British, Scandinavian, Russian, other Eastern European, Belgian/Dutch, Swiss/Austrian, French, Arab, German, Greek, Hungarian, Irish, Italian, Polish, Portuguese, or Other. Hoxby counts Canadians as British (but French Canadians as French), and includes sub-Saharan African, "USA," West Indian, and unclassified ancestry in "Other." I assign both Canadians and French Canadians to the "Other" category, and exclude sub-Saharan Africa, "USA," West India, and unclassified from the computation on the grounds that these groups are unlikely to be of white race. Hoxby's Hispanic index is the sum of the squared shares of the Hispanic population with Mexican, Puerto Rican, Cuban, other Central American, or South American ancestry. I add a sixth category, "Other Hispanic;" Hoxby folds this category into the South American group.

Summary statistics for each of these variables are reported in Table A3. All correlate highly across samples, with the only visible deviations appearing in the income variables.

c. Covariate construction, metropolitan level

Hoxby uses the Office of Management and Budget's June 30, 1990 definitions of Metropolitan Statistical Areas (http://www.census.gov/population/estimates/metro-city/90mfips.txt). In larger agglomerations, she uses PMSAs as the metropolitan construct in place of the larger CMSAs.

Hoxby writes, "I derive demographic measures at the metropolitan-area level from the *City and County Data Book*" (2000, p. 1221-2). In fact, her code derives MSA-level demographic characteristics by summing the SDDB observations for all districts attributed to the MSA. As noted previously, this introduces two problems. First, some areas are served by overlapping elementary and secondary school districts; Hoxby's approach double-counts residents of these areas toward metropolitan averages. Second, there are errors in the crosswalk file that Hoxby uses to assign districts to metropolitan areas. This can lead to serious misstatement of MSA totals. To take one extreme example, the SDDB contains records for many subdistricts within the single school district serving the state of Hawaii, while the CCD contains only a single record. As a result, only the administrative headquarters for the Hawaii district is assigned to Honolulu, and the Honolulu MSA is recorded as having only 13,700 residents in 0.77 square miles and. (The true values are 836,000 residents and 600 square miles.)

In place of the aggregated SDDB data, I use the Summary Tape File 3A of the 1990 Census to construct MSA demographics for both replication samples. I use county-level records for non-New England MSAs and town-level ("county subdivision") records for MSAs in New England. The source data for the STF-3A file are the same Census long form data as those from which the SDDB is constructed, but my approach avoids the imperfect match to MSAs and the double-counting problems that arise with the SDDB data.

- Metropolitan population and land area (*m_pop_n* and *m_arealand*): These are straightforward, but for the errors introduced when they are computed by summing school districts. The Hoxby/NELS data overstate the population by at least 10% in 71 MSAs, largely due to double-counting the populations of overlapping districts.
- Mean income (*ch_m_avglnmeaninc* and *m_lnmeanincA*): Though Hoxby's paper describes one of her control variables as "Mean of log(income) of metropolitan area," she does not make clear that this mean is taken over school districts rather than over individuals. That is, she computes the mean income (in levels) for each district, computes the log of this mean, then averages the district log(mean income) across districts in the MSA (weighting by the number of households) to form her MSA variable. This construction cannot be performed using the STF-3A data, and for this one MSA-level variable I follow Hoxby (for the close replication sample) in deriving it from the SDDB, after correctly assigning districts to MSAs. My preferred replication sample uses the more straightforward log of the MSA mean income.
- Gini coefficient (*ch_m_gini*, *jr_m_gini*): These are constructed identically to the district-level variables, as described above, using the STF-3A data in place of the SDDB.
- Racial composition of the MSA (*ch_m_pop_fas*, *ch_m_pop_fbl*, and *ch_m_pop_fhi*; *jr_m_pop_fas*, *jr_m_pop_fbl*, and *jr_m_pop_fhi*). Again, the MSA-level construction is the same as the district-level.

- Educational distribution (*m_ed_scol*, *m_ed_ba*, and *m_herfed*): Once more, the same as at the district level.
- Census division (*ch_d_div1-ch_d_div9*, *jr_m_divis1-jr_m_divis9*): Hoxby treats division as a district-level characteristic, computing dummy variables based on the division in which the district is located. She uses a somewhat different construction for her MSA-level first stage analysis, hand-coding each MSA that straddles divisions to the division in which the plurality of its population resides. My close replication sample follows Hoxby in each construction. Note, however, that my use of a repaired district-MSA crosswalk avoids the problems (discussed above) that appear in Hoxby's implementation. For my preferred sample, I treat division as an exclusively metropolitan-level characteristic, and I assign multi-division MSAs fractional values of the division indicators corresponding to the fraction of the MSA population in each division.

Summary statistics for each of these variables are reported in Table A4. They diverge more across samples than did individual or district-level variables, largely because of the differences between what is obtained from aggregated, overlapping SDDB data and from aggregated Census STF records that do not overlap.

4. Choice index

Hoxby's choice index is $c_m = 1 - \sum_{jm} \left(n_{jm} / N_m \right)^2$, where n_{jm} is the enrollment of district j in MSA m and N_m is the total enrollment in the MSA. Enrollment is drawn from the SDDB "Top 100" file, which in turn draws the variable from the CCD. For my replication samples, I use the 1989-90 CCD enrollment data directly.

Although I follow Hoxby's construction exactly for the close replication sample, I make a slight alteration for the preferred replication sample. Where Hoxby constructs her index using enrollment in all grades, I compute it considering only grade-8 enrollment. This makes little difference in MSAs with unified school districts. Where there are separate elementary and secondary districts—which cannot be said to compete against each other—Hoxby's formula will indicate more competition than actually exists (Urquiola, 1999).

5. Instruments

As instruments for the degree of choice in the MSA, Hoxby uses the number of larger and smaller rivers flowing through the MSA. She describes their construction as follows:

> The streams variables are derived from the U.S. Geological Survey's (USGS) 1/24,000 quadrangle maps. It was by using these extremely detailed maps—which allow the viewer to identify even very small streams, buildings, and boundaries—that I initially recognized the relationship between natural barriers and school district boundaries. The measurement of the streams variable was in two stages. Using the physical maps, I first counted all streams that were at least 3.5 miles long and of a certain width on the map. These data were checked against the Geological Survey's *Geographic Names Information System* (GNIS) for accuracy. I derived smaller streams directly from the GNIS. I employ two stream variables: the number of larger streams (measured by hand and often traversing multiple districts, sometimes multiple counties) and the number of smaller streams (from GNIS). (Hoxby, 2000, p. 1222).

Given the vagueness of Hoxby's description of "larger streams" and the subjectivity of the hand measurement, I opted not to try to reproduce Hoxby's counts. This has one unfortunate consequence: Hoxby does not provide counts for all MSAs, and as a result some MSAs must be excluded from analyses that include her larger streams instrument.

Hoxby's description of the genesis of the smaller streams measure is incomplete. She in fact measures *total* streams in the GNIS, and constructs smaller streams as total minus larger streams. As a result, any inaccuracies in the larger streams measure appear (with the opposite sign) in the smaller streams variable. Note, however, that the instrument set can be equivalently formulated as total and larger streams, as these span the same space as do Hoxby's measures.

The Hoxby/NCES program (2004b, lines 2811-2) cites the USGS web site as the source of her GNIS data, with a date of 2004. The GNIS is continually updated as better maps and surveys are completed. It is not clear what might be the impact of changes made between Hoxby's original analysis and her later extraction of the current GNIS data.

In the published paper, Hoxby writes that the GNIS "provides the longitude and latitude of [smaller streams'] origin and destination" (2000, note 24, p. 1222). Her code, however, does not make use of these variables, except to construct the streams' lengths. (She discards streams shorter than one mile.) Another variable, describing the county in which the stream's "destination" (i.e. mouth) is located, is used instead. The CCD is used to associate counties with MSAs, which means that some areas are mis-assigned.[14] Figure A2 indicates the implications of Hoxby's assignment algorithm for the Mississippi River. The Mississippi's mouth is in Plaquemines Parish, Louisiana, which is non-metropolitan. As a result, Hoxby's algorithm does not include the Mississippi in the total streams count of any of the eight MSAs through which it flows. Of course, the Mississippi is almost certainly included as a "larger" stream, but its exclusion from the total streams count means that smaller streams are necessarily undercounted.

When I reproduce the total streams variable for the replication sample, I use the same GNIS data as Hoxby, continuing to assign streams to counties on the basis of their destinations. I use an accurate county-MSA crosswalk, however, and compute population shares from Census STF data on non-overlapping towns ("county subdivisions"). Following Hoxby, I compute smaller streams by subtracting the number of larger streams that Hoxby counted from the number of total streams that I obtain. This produces a negative number in five MSAs (two in New England), where Hoxby evidently counted more larger streams than the number of total streams (stream mouths) than the GNIS destination variable indicates.[15]

The alternative instruments discussed in Section III use a more expansive definition of "total streams," in which a stream is counted toward any MSA through which it flows, regardless of where it terminates. For this purpose, I use an alternative version of the GNIS database (Geographic Names Office, 1999), which contains a field listing all counties through which each stream flows. (For the Mississippi, this list contains 117 counties, which are shown on Appendix Figure A2.) In New England, where some counties are split into non-metropolitan and metropolitan components, I compute weights for each county

---

[14] When counties are not wholly contained within a single MSA (in New England, and elsewhere when districts are mis-assigned), all streams in each county are assigned to the MSA in which the plurality of the county's population resides. The population shares for this computation are taken from the SDDB, so double-count some areas served by multiple districts.

[15] Note that there are no negative "smaller streams" counts in the Hoxby/NCES data; the problem arises only when I re-create Hoxby's total streams variable with a corrected county-MSA crosswalk in place of Hoxby's imperfect district-MSA crosswalk.

corresponding to the share of the population in the county residing in each MSA. Using the county population weights, streams are partially assigned to each MSA with which a county intersects.[16] Outside of New England, I need not approximate MSA boundaries, and each stream gets a weight of one toward each MSA through which it flows. My "total streams" count is the sum of the weights of all streams flowing through each MSA.

I consider three categorizations of the resulting total streams count into larger and smaller streams. First, I use Hoxby's larger streams variable, defining smaller streams as my total streams count minus Hoxby's larger streams for those MSAs for which Hoxby's variable is available. There are six MSAs, five in New England, for which this indicates a negative number of smaller streams.[17] Second, I compute separate counts of inter-county and intra-county streams, where an inter-county stream is one that flows through more than one county. Finally, using the latitude and longitude variables (which exist on both versions of the GNIS data), I compute the distance between the source and mouth of each stream, and compute separate counts of streams that are longer than and shorter than 3.5 miles.[18]

Table A5 presents summary statistics for the various streams variables

6. Weights

Hoxby writes that her student-level regressions are "weighted so that each metropolitan area receives equal weight," (2000, p. 1226). This is operationalized by summing the NELS sample weights for all students in each MSA, then dividing each individual's weight by the MSA's total. In the Hoxby/NCES data, however, this normalization is carried out once for each survey wave, without regard to missing values. Many observations that are excluded from the regression analyses because they are missing a test score or one of the covariates are nevertheless included in the weight normalization. As a result, the weights sum to substantially less than one in each MSA when the sum is taken only over observations that appear in the regressions. In the replication samples, I re-

---

[16] For streams flowing through multiple counties in the same MSA, the maximum population weight is used. (This is analogous to assigning a weight of one to any stream flowing through any metropolitan county in the remainder of the country.)

[17] The non-New England MSA is Topeka, Kansas. Hoxby reports 82 larger streams, but I count only 36 stream mouths and 41 total streams. Hoxby's count of 109 total streams appears to derive from the CCD's assignment of Jefferson County and Osage County school districts to the Topeka MSA. These counties were part of the MSA in 1981, but have been excluded since the release of the 1983 definitions.

[18] There are a small number of streams missing latitude and longitude information, for which length cannot be computed. These are included in earlier counts, but are excluded from the categorization by length.

normalize for each specification, including only observations used in that specification. Note (Table A2) that the average weight is substantially smaller in the Hoxby/NCES sample than in the replication samples, and that the product of the average weight and the number of observations in the analysis samples equals the number of MSAs for the replication samples but not in the Hoxby/NCES sample.

Appendix B: Econometric Specification and Standard Error Computation

Simplifying notation slightly, Hoxby specifies her model for student achievement as

(B1) $$A_{ikm} = C_m\beta_1 + X_{ikm}\beta_2 + \overline{X}_{km}\beta_3 + \overline{X}_m\beta_4 + e_{ikm},$$

where A_{ikm} is the test score of student i in district k in MSA m; C_m is the choice index for MSA m; and X_{ikm}, $\overline{X}_{km}$, and $\overline{X}_m$ are individual, district mean, and MSA mean covariates.[19] As there are likely omitted variables at all three levels, it is unreasonable to assume that e_{ikm} is independent across individuals within the same district or MSA. Hoxby specifies an error components model,

(B2) $$e_{ikm} = \varepsilon_m + \varepsilon_{km} + \varepsilon_{ikm},$$

assuming that each component is independent and identically distributed across observations at that level. That is,

(B3) $$\text{cov}\left(e_{ikm}, e_{jdn}\right) = \begin{cases} 0 & \text{if} \quad m \neq n \\ \sigma_m^2 & \text{if} \quad m = n, k \neq d \\ \sigma_m^2 + \sigma_k^2 & \text{if} \quad m = n, k = d, i \neq j \\ \sigma_m^2 + \sigma_k^2 + \sigma_i^2 & \text{if} \quad m = n, k = d, i = j \end{cases}$$

where $\sigma_m^2 = \text{var}\left(\varepsilon_m\right)$, $\sigma_k^2 = \text{var}\left(\varepsilon_{km}\right)$, and $\sigma_i^2 = \text{var}\left(\varepsilon_{ikm}\right)$.

[19] Hoxby (2000) devotes considerable discussion to the complication that the district heterogeneity variables do not average to their MSA-level analogues. I neglect this complication; one may simply think that $\overline{X}_{km}$ and $\overline{X}_m$ are arbitrary vectors of district- and MSA-level covariates. One related point is worth mentioning, however. In her footnote 21, Hoxby claims that when the district heterogeneity variables are excluded, the individual-level first stage model (analogous to the outcome equation indicated above) is identical to an MSA-level analogue that excludes X_{ikm} and $\overline{X}_{km}$, as there can be no correlation between these two vectors and the MSA-level choice variable conditional on $\overline{X}_m$. This is correct only so long as the sample average of the individual and district-level variables is identical to the MSA means. When, as in this case, the individual variables are observed only for a sample, this will not be true, and there is no guarantee that coefficients on MSA-level covariates are invariant to the level at which the model is estimated.

It is useful to work with matrix notation. Let $W = [C_m \ X_{ikm} \ \overline{X}_{km} \ \overline{X}_{ikm}]$ be the matrix of right-hand-side variables from (B1), and $\beta = [\beta_1' \ \beta_2' \ \beta_3' \ \beta_4']'$. (B1) then becomes $A = W\beta + e$. If C_m is endogenous, $E[W'e] \neq 0$. Let R_m be the streams instrumental variables, and let $Z = [R_m \ X_{ikm} \ \overline{X}_{km} \ \overline{X}_{ikm}]$. If the instruments are valid, $E[Z'e] = 0$, and

(B4) $$\hat{\beta}_{IV} = (W'P_Z W)^{-1} W'P_Z A = \beta + (W'P_Z W)^{-1} W'P_Z e,$$

where $P_Z = Z(Z'Z)^{-1}Z'$.

Inference proceeds by noting that

(B5) $$\text{var}(\hat{\beta}_{IV}) = (W'P_Z W)^{-1} W'P_Z' \Gamma P_Z W (W'P_Z W)^{-1},$$

where $\Gamma = E[ee']$ is the variance-covariance matrix with elements specified in (B3). Under conventional assumptions of i.i.d. observations (i.e. with $\sigma_m^2 = \sigma_k^2 = 0$), $\Gamma = \sigma_i^2 I$, and we obtain the conventional variance expression, $\text{var}(\hat{\beta}_{IV}) = \sigma_i^2 (W'P_Z W)^{-1}$. This expression does not apply, however, with non-zero district- and MSA-level error components.

Hoxby's so-called "Moulton" formula for $\text{var}(\hat{\beta}_{IV})$ (developed by Moulton, 1986, for the OLS case) simply forms a consistent estimate of Γ from estimates of the error component variances, σ_m^2, σ_k^2, and σ_i^2. The resulting $\hat{\Gamma}$ is then plugged in for Γ in (B5) to estimate $\text{var}(\hat{\beta}_{IV})$. This approach requires only that the three error component variances be estimated consistently. Several consistent estimators are available, and it is not clear that any one should be preferred to another. The options multiply in the "unbalanced panel" case that is relevant here, where districts and MSAs contain unequal numbers of observations. Hoxby has not provided code for her implementation, and does not specify how she estimates the variance parameters.

My implementation of the Moulton approach estimates the error component variances from contrasts between individual, district-mean, and MSA-mean residual variances, extending Greene's (2000, p. 570-2) discussion of random effects in unbalanced panels to the three-component hierarchy considered here. Let $\bar{e}_{km} = \frac{1}{N_{km}} e_{ikm}$ and $\bar{e}_m = \frac{1}{N_m} e_{ikm}$ be the district- and MSA-level means of e_{ikm}, where N_{km} is the number of individual observations at district *k* in MSA *m* and $N_m = \sum_k N_{km}$ is the number of individuals in the MSA. Ignoring degrees of freedom corrections, which only complicate the notation,

(B6a) $$\text{var}\left(e_{ikm} - \bar{e}_{km}\right) = \text{var}\left(\varepsilon_{ikm}\right) = \sigma_i^2 ;$$

(B6b) $$\text{var}\left(\bar{e}_{km} - \bar{e}_m\right) = \text{var}\left(\varepsilon_{km} + \frac{1}{N_{km}}\varepsilon_{ikm}\right) = \sigma_k^2 + \frac{1}{N_{km}}\sigma_i^2 \text{; and}$$

(B6c) $$\text{var}\left(\bar{e}_m\right) = \text{var}\left(\varepsilon_m + \frac{1}{P_m}\varepsilon_{km} + \frac{1}{N_m}\varepsilon_{ikm}\right) = \sigma_m^2 + \frac{1}{P_m}\sigma_k^2 + \frac{1}{N_m}\sigma_i^2 ,$$

where P_m is the number of districts in MSA *m*. Note that with unbalanced data, N_{km} can vary across districts, as can P_m and N_m across MSAs. Solving the sample analogues of (B6a), (B6b), and (B6c) for the variance components thus involves the sample averages of N_{km}^{-1}, N_m^{-1}, and P_m^{-1}, but these are readily estimated from the data. Code for my implementation may be downloaded from my web site.

I also present so-called "clustered" standard error estimates. These proceed from the observation that we do not require a consistent estimate of Γ, but only of $Z'\Gamma Z$. This can be consistently estimated under substantially weaker assumptions than (B3). In particular, we can allow for arbitrary heteroskedasticity and within-MSA correlation patterns. Let Z^m be the sub-matrix of *Z* corresponding to MSA *m*, and let e^m be the analogous subvector of *e*. So long as $\text{cov}\left(e_{ikm}, e_{jdn}\right) = 0$ whenever $m \neq n$, $G = \sum_m Z^{m}{}' e^m e^{m}{}' Z^m$ is consistent for $Z'\Gamma Z$. The cluster variance estimator, then, is

(B7) $$\text{var}\left(\hat{\beta}_{IV}\right) = \left(W' P_Z W\right)^{-1} W' Z\left(Z'Z\right)^{-1} G\left(Z'Z\right)^{-1} Z'W\left(W' P_Z W\right)^{-1}.$$

As with the Moulton estimator, asymptotic consistency is achieved as the number of MSAs goes to infinity. An important reason to prefer the "cluster" estimator is that it is fully automated, where the Moulton estimator requires the researcher to choose among several estimators of the error component variances and is therefore more difficult to replicate.

1. Exogeneity and overidentification tests

It is useful to test for the endogeneity of the choice index, C_m. With i.i.d. errors, one might use a Hausman test: Under the null hypothesis that C_m is exogenous, the OLS estimator of β is efficient but the IV estimator is consistent; under the alternative of endogeneity, OLS is inconsistent but IV remains consistent. With the serial correlation implied by the error components model, however, OLS is no longer efficient even under the null hypothesis. Thus, the exogeneity tests reported in the text use an alternative, "artificial regression" test that can be made consistent to clustering (Davidson and MacKinnon, 1993, p. 237-42). I form $\hat{C}_m$, the fitted values of C_m from the first stage regression—using the same sample used for IV—and estimate the regression A_{ikm} on C_m, $\hat{C}_m$, X_{ikm}, $\overline{X}_{km}$, and $\overline{X}_m$. Under the null hypothesis of exogeneity, the coefficient on $\hat{C}_m$ should be zero; under the alternative, it should be non-zero. I report tests of the significance of this coefficient that use clustered standard errors, although one might equally well use the Moulton standard errors for this purpose (and indeed they give similar results).

In specifications involving multiple instruments, one might also like to report overidentification tests of the mutual consistency of the different instruments. Again, conventional Hausman test-based approaches do not work with serially correlated errors, as the regular two-stage least squares estimator is not efficient. Hoxby and Paserman (1998) propose a Hausman test for the error components model that is based on the GMM estimate of the overidentified specification. Following Hoxby (2000), I present only 2SLS coefficient estimates and not the more efficient GMM estimates. Accordingly, I choose not to present GMM-based overidentification tests, but instead simply report specifications that omit the suspect instruments.[20]

Appendix C: Full coefficient vectors

Tables C1 through C5 present the full coefficient vectors from the models in Tables 1 through 5 of the main text.

Appendix D: Additional Specifications

The remaining appendix tables present extensions and alternative specifications. Table D1 presents estimates of the first-stage regression (analogous to those in Table 3) using the "preferred" replication sample. As before, Hoxby's larger streams variable plays a substantially different role in the MSA-level "implied first stage" than in the actual first stages to the individual-level models. Table D2 presents the corresponding IV estimates

[20] A second reason to avoid overidentification tests in this context is that they are likely to have little power when, as here, only one of the instruments plays an important role in the first stage.

from the preferred replication sample. As in the "close" sample (Table 4), only specifications that include Hoxby's larger streams variable as an instrument show any indication that choice is endogenous or that it has a significant positive effect on outcomes. OLS estimates are negative (though far from significant) for both 8th and 12th grade reading scores.

Table D3 extends the sample selection analysis from Table 5 to 8th grade reading scores. Choice coefficients fall somewhat when district-level covariates are dropped from Hoxby's basic specification, and fall substantially farther when public schools are assigned to MSAs on the basis of their zip codes. There is an additional downward effect of adding private schools to the sample, and *t*-statistics are uniformly below one in this specification.

Table D4 extends the analysis to the four NELS test scores that have not yet been considered. There is no indication that choice has a different effect on mathematics scores or on 10th grade reading scores than it does on the 8th and 12th grade reading scores considered thus far.

Table D5 presents a version of the MSA-level first stage estimated only on MSAs in the NELS sample. Comparing this to Table 3, it is clear that the divergence between Hoxby's MSA-level first stage (including all MSAs) and the individual-level MSAs derives from the sample of MSAs included rather than from the level at which the model is estimated.

Table D6 presents "two sample" IV estimates of the basic model, in which the first stage is estimated at the MSA level, predicted values are formed, and these predicted values are used in the individual-level second stage regression. In each panel, the first row presents estimates that use only NELS MSAs for the first stage, while the second row presents estimates in which the full set of MSAs are used to estimate the first stage model. This approach introduces two complications. First, the individual- and district-level covariates must be excluded from the MSA-level first stages. This does not affect the IV estimates of the choice coefficient, so long as all first-stage controls are included in the second stage as well, but changes the interpretation of the second-stage coefficients on the control variables (which are not shown in any case). Second, special formulas are needed to compute correct standard errors. I do not compute these, but merely present clustered S.E.s from the second stage regression; these will tend to understate the true variability of the indicated coefficients.

As with conventional IV, the two-sample estimates indicate significant choice effects when Hoxby's instruments are used and when only the NELS MSAs are included in the first stage. When the first stage sample is broadened to include all MSAs, the choice coefficient falls in the specifications using Hoxby's instruments but not in those using inter- and intra-county streams as instruments. Even with the incorrect standard errors, none of the specifications indicates a significant choice effect when all MSAs are used to compute the first stage.

Table D7 presents estimates of the reduced form relationship between streams and student test scores. These are quite noisy, and the streams coefficients are never significantly different from zero in models that exclude Hoxby's larger streams variable. In models including this variable, it always has a negative coefficient, while smaller streams have a significantly positive coefficient. When these variables are added together, however, the "total streams" coefficient is statistically and substantively indistinguishable from zero.

Finally, Table D8 presents IV estimates for MSA average test scores, computed entirely at the MSA level. Very few of these are significant, even when averages of individual- and district-level covariates are included in the specification (in which case the specification is essentially identical to Hoxby's, but is more conservative about the standard error computation). The two models that are significant both use Hoxby's streams variables.

Appendix E: Estimates of the choice effect on spending

Much of Hoxby's discussion, and all of my analysis thus far, focuses on the relationship between district fragmentation and student achievement. However, her conclusions relate to the choice effect on school productivity, and she presents analyses of the relationship between choice and school spending in her Tables 5 and 6. The relevant coefficients from her preferred specification are reproduced in Panel A of Table E1. Panel B reproduces this specification in the Hoxby/NCES data. Standard errors are much larger here, and indeed the published standard error from the IV specification much more closely resembles the Hoxby/NCES classical standard error—which assumes that observations are i.i.d.—than it does either the Moulton or clustered errors from that sample. There is no indication in the Hoxby/NCES sample of a significant IV effect of choice on per-pupil spending, though the OLS estimate is weakly significant. Panels C and D repeat the same specification in the replication samples, and also present estimates that use inter- and intra-county streams as instruments. Again, there is no indication of a significant effect in IV, but OLS estimates are significantly negative.

Hoxby's model, however, may be misspecified. She estimates it at the school district level, and takes as her dependent variable the log of per pupil spending in the district. The convexity of the log transformation is the source of the problem. Consider an MSA with two equally-sized schools, A and B, with different levels of per-pupil spending, y_A and y_B. If these schools are in the same district, Hoxby's dependent variable will be $\ln(0.5*(y_A + y_B))$. If the schools are divided into two districts, however, the average of Hoxby's dependent variable in the MSA will be $0.5*(\ln(y_A) + \ln(y_B))$, which is smaller. Thus, without any behavioral effect of choice at all, one would estimate a negative coefficient in Hoxby's regression.

One way to avoid any bias that this mechanical relationship might introduce is to estimate the specification at the MSA level, taking as the dependent variable the log of average per pupil spending in the MSA. This is done in columns 4-6 of Table E1. Evidently, the mechanical bias is not an issue—estimates are nearly identical (with very similar standard errors) to those obtained at the district level).

References

Alesina, Alberto F.; Baqir, Reza and Hoxby, Caroline M. "Ethnic Diversity and the Number of Jurisdictions." Working paper, Harvard University, August 1999.

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

ESRI. "Data & Maps 2002." CD, ESRI, 2002.

Geographic Names Office. "National Geographic Names Database." GNIS Digital Gazeteer CD, U.S. Geological Survey, 1999.

Greene, William H. *Econometric Analysis*. New Jersey: Prentice Hall, 2000.

Hoxby, Caroline M. "Does Competition among Public Schools Benefit Students and Taxpayers?" *American Economic Review*, December 2000, *90*(5), pp. 1209-38.

____. "District-Level and Metropolitan-Area Variables Merged with NELS Data." CD, National Center for Education Statistics, September 2, 2004a.

____. "Documentation to 'District-Level and Metropolitan-Area Variables Merged with NELS Data': Construct.Do." National Center for Education Statistics, September 2, 2004b.

____. "Documentation to 'District-Level and Metropolitan-Area Variables Merged with NELS Data': Runregressions.Do." National Center for Education Statistics, September 2, 2004c.

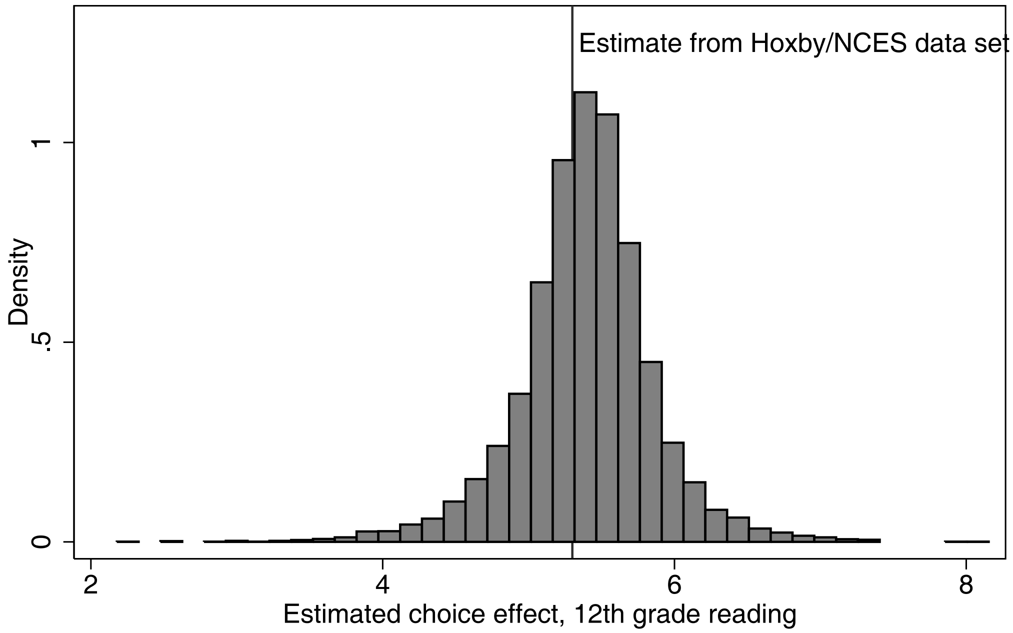
Hoxby, Caroline M. and Paserman, M. Daniele. "Overidentification Tests with Grouped Data." National Bureau of Economic Research Working paper #T0223, February 1998.

Moulton, Brent R. "Random Group Effects and the Precision of Regression Estimates." *Journal of Econometrics*, August 1986, *32*(3), pp. 385-97.

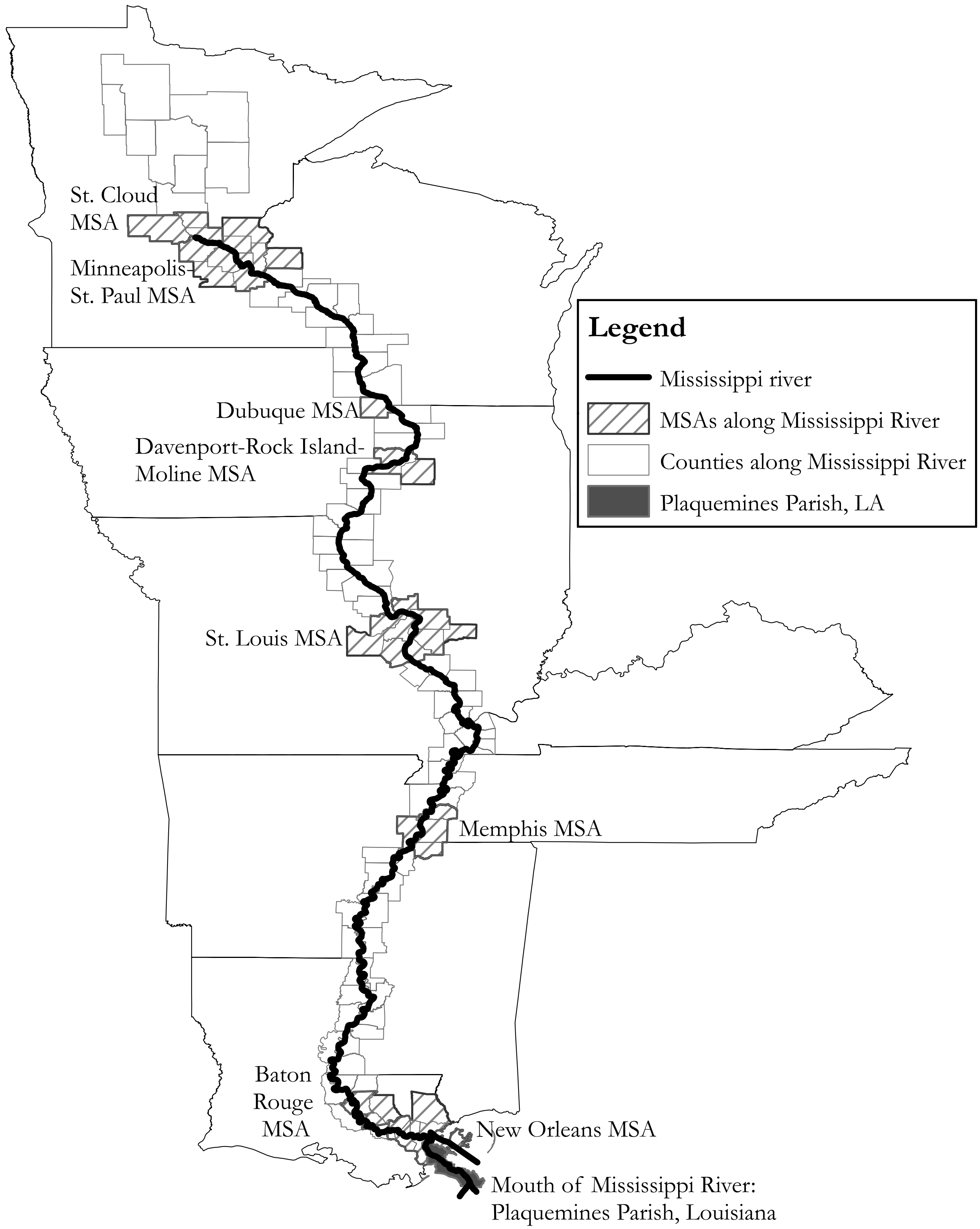
StataCorp. *Stata Base Reference Manual*. College Station, Texas: Stata Press, 2003.

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

Appendix Figure A1. Distribution of estimated choice effect across replications of Hoxby/NCES data assembly program



Appendix Figure A2. Alternative algorithms for assigning the Mississippi River to MSAs



Appendix Table A1. Segment of Hoxby's construct.do program (Hoxby, 2004b) that merges NELS student file to NELS school file.

```
*Infix data from stmeg.pri and scmeg.pri.;

cd nels;
infix using nels.fromstmeg.infix.dct, using(d:\nels92\stmeg.pri);
sort sch_id;
compress;
save nels.fromstmeg.dta, replace;
clear;
infix using nels.fromscmeg.infix.dct, using(d:\nels92\scmeg.pri);
sort sch_id;
compress;
save nels.fromscmeg.dta, replace;
clear;

*Merge the variables from the two datasets above.;

use nels.fromstmeg.dta;
merge sch_id using nels.fromscmeg.dta;
tab _merge;
drop if _merge==2;
drop _merge;
save nels.dta, replace;
use nels.fromscmeg.dta;
sort f1sch_id;
save nels.fromscmeg.dta, replace;
use nels.dta;
sort f1sch_id;
merge f1sch_id using nels.fromscmeg.dta;
tab _merge;
drop if _merge==2;
drop _merge;
save nels.dta, replace;
use nels.fromscmeg.dta;
sort f2sch_id;
save nels.fromscmeg.dta, replace;
use nels.dta;
sort f2sch_id;
merge f2sch_id using nels.fromscmeg.dta;
tab _merge;
drop if _merge==2;
drop _merge;
save nels.dta, replace;
```

Appendix Table A2: Summary statistics for individual-level covariates, NELS

	Hoxby/NELS data (1)		Close replication (2)		Preferred replication (3)		Correlations		
	Mean	S.D.	Mean	S.D.	Mean	S.D.	(1) & (2)	(1) & (3)	(2) & (3)
Public school (BY)	0.80	(0.40)	0.80	(0.40)	same as close		1.000		
Public school (F2)	0.87	(0.33)	0.87	(0.33)	same as close		1.000		
ln(fam. Income)	3.36	(0.94)	3.36	(0.94)	10.19	(1.21)	1.000	0.999	0.999
Asian	0.06	(0.24)	0.06	(0.24)	0.07	(0.25)	1.000	0.996	0.996
Hispanic	0.13	(0.34)	0.13	(0.34)	0.14	(0.35)	1.000	0.999	0.999
Black	0.12	(0.33)	0.12	(0.33)	0.13	(0.33)	1.000	0.996	0.996
Female	0.50	(0.50)	0.50	(0.50)	0.49	(0.50)	1.000	1.000	1.000
Parents some college	0.23	(0.42)	0.23	(0.42)	0.38	(0.49)	1.000	0.453	0.453
Parents BA+	0.37	(0.48)	0.37	(0.48)	0.30	(0.46)	1.000	0.759	0.759
8th grade reading score	50.5	(10.2)	same as Hoxby/NELS						
12th grade reading score	50.7	(10.0)	same as Hoxby/NELS						
8th gr. weight (*100)	0.74	(1.26)	1.78	(2.04)	1.57	(1.48)	0.743	0.963	0.794
12th gr. weight (*100)	0.79	(3.01)	3.27	(4.97)	2.98	(5.31)	0.640	0.413	0.825
Fr. in MSA									
8th grade	0.56		0.70		0.55				
12th grade	0.53		0.66		0.52				
Analysis sample size									
8th grade reading	10,175		10,429		11,719				
12th grade reading	5,475		5,934		6,688				
Fraction of district assignments that agree									
8th grade							0.65	0.73	0.75
12th grade							0.65	0.41	0.59
Fraction of MSA assignments that agree									
8th grade							0.77	0.94	0.81
12th grade							0.77	0.68	0.66

Notes: All statistics are unweighted. See text for details. N=27,805, though many variables are missing for many observations.

Appendix Table A3: Summary statistics for district-level covariates, SDDB

	Hoxby/NELS data		Close replication		Preferred replication		Correlations		
	(1)		(2)		(3)		(1) & (2)	(1) & (3)	(2) & (3)
	Mean	S.D.	Mean	S.D.	Mean	S.D.			
Fr. Asian	0.02	(0.08)	0.01	(0.02)	0.01	(0.02)	1.000	1.000	1.000
Fr. Hispanic	0.05	(0.12)	0.05	(0.12)	0.05	(0.12)	1.000	1.000	1.000
Fr. Black	0.04	(0.11)	0.04	(0.11)	0.04	(0.11)	1.000	1.000	1.000
Racial homog. index	0.84	(0.17)	0.85	(0.17)	0.85	(0.17)	1.000	1.000	1.000
Fr. some college	0.25	(0.08)	same as Hoxby/NELS						
Fr. BA+	0.15	(0.10)	same as Hoxby/NELS						
Educ. homog. index	0.31	(0.05)	same as Hoxby/NELS						
log(mean HH inc.)	10.37	(0.32)	10.37	(0.34)	same as close		0.986		
Gini coeff.	0.38	(0.05)	0.38	(0.05)	0.40	(0.06)	0.993	0.928	0.934
Ethnic homog. index	0.78	(0.21)	0.79	(0.21)	0.81	(0.21)	1.000	1.000	1.000
Fr. in MSAs	0.39		0.40		same as close				
Fraction of MSA assignments that agree							0.97	0.97	1.00

Notes: All statistics are unweighted. See text for details. N=14,947 in replication samples, 15,304 in Hoxby data.

Appendix Table A4: Summary statistics for MSA-level covariates, SDDB

	Hoxby/NELS		Close		Preferred		Correlations		
	(1)		(2)		(3)		(1) & (2)	(1) & (3)	(2) & (3)
	Mean	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**			
Choice	0.67	(0.27)	0.66	(0.27)	0.66	(0.27)	0.971	0.966	0.995
Population (10 millions)	0.063	(0.112)	0.058	(0.098)	same as close		0.985		
ln(Population)	12.66	(1.12)	12.63	(1.00)	same as close		0.960		
Land Area (100,000s of sq. miles)	0.017	(0.019)	0.017	(0.020)	same as close		0.958		
ln(Land Area)	7.06	(1.01)	7.07	(0.89)	same as close		0.871		
Fr. Asian	0.02	(0.05)	0.02	(0.04)	0.02	(0.04)	0.992	0.992	1.000
Fr. Hispanic	0.07	(0.13)	0.07	(0.13)	0.07	(0.13)	0.999	0.999	1.000
Fr. Black	0.09	(0.09)	0.10	(0.09)	0.10	(0.09)	0.988	0.988	1.000
Racial homog. index	0.72	(0.16)	0.72	(0.16)	0.71	(0.16)	0.992	0.991	1.000
Fr. some college	0.28	(0.05)	0.26	(0.05)	same as close		0.893		
Fr. BA+	0.19	(0.06)	0.20	(0.07)	same as close		0.980		
Educ. homog. index	0.27	(0.02)	0.27	(0.02)	same as close		0.954		
log(mean HH inc.)	3.56	(0.17)	10.48	(0.18)	10.50	(0.19)	0.987	0.976	0.992
Gini coeff.	0.40	(0.02)	0.40	(0.02)	0.43	(0.03)	0.963	0.906	0.937
Ethnic homog. index	0.64	(0.19)	0.63	(0.19)	0.63	(0.19)	0.991	0.991	1.000
# of invalid MSA codes	5								

Notes: N=327 MSAs in Hoxby data, 335 in replication samples. All analyses are unweighted

Table A5: Summary statistics and correlations for streams variables

	Mean	S.D.	Correlation matrix				
			(1)	**(2)**	**(3)**	**(4)**	**(5)**
Larger streams							
Published	*7.9*	*(14.8)*					
(1) Hoxby/NCES data (N=314)	44.3	(64.1)	1.00				
(2) Inter-county streams	41.2	(33.7)	0.50	1.00			
(3) Long streams (>3.5 miles)	69.9	(57.8)	0.55	0.84	1.00		
Total streams							
(1) Hoxby/NCES data (N=314)	128.3	(119.8)	1.00				
(2) Replication stream mouths (N=319)	124.4	(119.2)	0.97	1.00			
(3) Replication total streams	147.8	(149.2)	0.89	0.92	1.00		
Smaller streams							
Published	*182.7*	*(208.8)*					
(1) Hoxby/NCES data (N=314)	84.0	(78.0)	1.00				
(2) Replic. total mouths - Hoxby larger (N=304)	80.3	(82.6)	0.94	1.00			
(3) Replic. total streams - Hoxby larger (N=311)	108.1	(118.1)	0.82	0.86	1.00		
(4) Intra-county streams	106.6	(122.4)	0.78	0.78	0.90	1.00	
(5) Short streams (<3.5 miles)	75.3	(102.6)	0.73	0.73	0.86	0.96	1.00

Notes: N=335 except where otherwise indicated. All statistics are unweighted.

Appendix Table C1: Full coefficient vectors for models in Table 1

	12th grade reading score								8th grade reading score					
	Published		Hoxby/ NCES		Close replication		Preferred replication		Hoxby/ NCES		Close replication		Preferred replication	
	(1)		(2)		(3)		(4)		(2)		(3)		(4)	
Choice index	5.77		5.30		4.74		3.29		4.45		5.93		2.93	
	(2.21)		(2.36)		(1.98)		(1.83)		(1.87)		(2.10)		(1.58)	
Individual covariates														
ln(Family income)	1.54		1.62		1.75		0.85		1.60		1.67		1.03	
	(0.16)		(0.18)		(0.17)		(0.12)		(0.12)		(0.12)		(0.08)	
Asian	0.28		0.95		0.35		-1.47		0.25		0.17		-0.67	
	(0.59)		(0.51)		(0.49)		(0.43)		(0.38)		(0.38)		(0.34)	
Hispanic	-2.87		-1.71		-2.44		-3.49		-2.23		-2.58		-2.41	
	(0.52)		(0.47)		(0.45)		(0.41)		(0.32)		(0.32)		(0.29)	
Black	-5.49		-4.07		-4.58		-5.82		-4.14		-4.46		-4.03	
	(0.50)		(0.52)		(0.50)		(0.45)		(0.34)		(0.34)		(0.30)	
Female	1.96		2.31		2.18		2.00		2.25		2.27		2.25	
	(0.23)		(0.25)		(0.24)		(0.22)		(0.18)		(0.18)		(0.17)	
Parents' highest ed is BA+	5.45		5.07		5.14		6.51		5.42		5.13		6.98	
	(0.30)		(0.33)		(0.31)		(0.34)		(0.24)		(0.24)		(0.26)	
Parents' highest ed is some college	2.31		2.98		2.90		2.52		3.09		3.25		2.49	
	(0.30)		(0.33)		(0.32)		(0.28)		(0.23)		(0.23)		(0.20)	
District/MSA covariates	*Dist.*	*MSA*	*Dist.*	*MSA*	*Dist.*	*MSA*	*Dist.*	*MSA*	*Dist.*	*MSA*	*Dist.*	*MSA*	*Dist.*	*MSA*
Population (tens of millions)		nr		-1.59		-2.85		-0.10		-0.11		-1.08		0.00
				(1.09)		(1.38)		(1.29)		(0.84)		(1.35)		(1.03)
Land area (hundreds of thousands of sq. mi.)		nr		-13.13		-8.40		2.15		-14.68		-16.55		-6.35
				(7.66)		(5.85)		(6.25)		(6.12)		(6.02)		(4.63)
ln(mean HH income)	nr	-5.42	3.45	-3.66	1.83	-0.52	2.19	-1.84	5.08	-7.41	4.11	-4.73	3.87	-3.10
		(5.53)	(2.86)	(3.95)	(1.99)	(3.15)	(1.56)	(2.61)	(2.06)	(3.02)	(1.68)	(3.00)	(1.28)	(2.23)
Gini coefficient, HH income	nr	-12.77	8.55	-19.12	5.38	8.27	-6.32	5.79	6.85	-15.70	3.41	-4.05	0.63	-3.21
		(12.02)	(10.67)	(21.10)	(8.60)	(19.34)	(5.51)	(13.44)	(7.54)	(17.17)	(7.25)	(19.81)	(4.29)	(11.85)
Pct. Asian	nr	-5.62	6.42	-3.94	30.26	-39.41	8.52	-7.49	10.72	-16.60	16.07	-25.76	5.31	-12.85
		(13.07)	(6.73)	(8.18)	(6.77)	(10.72)	(5.47)	(8.96)	(4.73)	(6.01)	(4.89)	(8.67)	(4.34)	(7.52)
Pct. Black	nr	-0.73	14.38	-9.84	16.08	-14.61	6.46	1.98	-0.59	-3.39	4.24	-6.35	0.17	-3.65
		(6.06)	(3.39)	(7.25)	(3.40)	(7.23)	(2.93)	(6.31)	(2.13)	(5.28)	(2.36)	(6.57)	(2.00)	(4.99)
Pct. Hispanic	nr	0.25	10.84	-7.93	11.62	-8.88	7.18	1.09	-2.21	2.52	1.05	0.70	0.20	0.52
		(3.52)	(3.75)	(4.87)	(3.56)	(5.14)	(3.02)	(4.60)	(2.68)	(3.67)	(2.83)	(4.70)	(2.46)	(3.85)
Racial homogeneity index (0=homog., 1=heterog.)	nr	-9.60	87.42	-42.61	90.55	-51.73	52.72	-0.12	8.13	-8.13	30.07	-35.03	6.27	-13.43
		(7.84)	(17.46)	(20.51)	(17.99)	(28.31)	(15.99)	(26.77)	(11.72)	(15.71)	(13.67)	(28.20)	(11.87)	(22.17)
Ethnic homogeneity index	nr	16.31	-63.75	30.27	-66.22	35.65	-43.63	6.12	-6.87	6.81	-21.48	27.03	-3.65	11.38
		(10.70)	(13.36)	(15.88)	(13.64)	(21.53)	(12.17)	(20.34)	(9.10)	(12.22)	(10.59)	(21.54)	(9.17)	(16.95)
Pct. Some college	nr	5.27	1.42	-2.47	-4.04	16.72	4.19	2.39	-3.57	2.75	-9.25	18.75	-4.37	3.01
		(7.14)	(7.34)	(10.34)	(7.01)	(11.12)	(6.24)	(10.11)	(5.27)	(8.16)	(5.81)	(10.73)	(4.91)	(8.57)
Pct. College graduates	nr	3.16	1.39	2.53	1.11	3.58	1.62	4.59	-1.96	7.96	-0.48	6.59	-0.96	3.83
		(5.93)	(5.68)	(8.50)	(4.83)	(6.44)	(4.06)	(5.48)	(4.17)	(6.47)	(4.17)	(6.40)	(3.21)	(4.78)
Educational homogeneity index	nr	-5.45	-11.50	-8.59	-15.34	12.53	-15.67	-9.45	-9.18	-1.65	-17.01	11.38	-14.41	2.88
		(12.95)	(8.98)	(21.75)	(8.25)	(20.62)	(7.18)	(19.04)	(6.55)	(17.53)	(7.03)	(20.85)	(5.70)	(16.36)
N	6,119		5,475		5,934		6,688		10,175		10,429		11,719	
R^2	nr		0.161		0.173		0.200		0.181		0.180		0.197	

Notes: All models include census division fixed effects. Moulton S.E.s are reported. "nr"=not reported.

Appendix Table C2: Full coefficient vectors for first stage models in Table 2

	MSA level				Indiv. level (12th gr. reading samp.)					
	Published	Hoxby/ NCES	Close replication	Preferred replication	Hoxby/ NCES		Close replication		Preferred replication	
	(1)	(2)	(3)	(4)	(2)		(3)		(4)	
Larger streams (100s)	0.080	0.012	0.040	0.043	-0.043		-0.024		0.015	
	(0.040)	(0.021)	(0.021)	(0.021)	(0.023)		(0.020)		(0.020)	
Smaller streams (100s)	0.034	0.096	0.093	0.091	0.133		0.133		0.114	
	(0.007)	(0.019)	(0.018)	(0.018)	(0.021)		(0.017)		(0.018)	
District/MSA covariates	*MSA*	*MSA*	*MSA*	*MSA*	*Dist.*	*MSA*	*Dist.*	*MSA*	*Dist.*	*MSA*
Population (tens of millions)	150.00	0.09	0.07	0.06		0.11		0.03		-0.04
	(130.00)	(0.14)	(0.16)	(0.16)		(0.14)		(0.16)		(0.16)
Land area (hundreds of thousands of sq. mi.)	0.50	1.43	0.90	1.00		0.53		0.65		0.60
	(0.50)	(0.79)	(0.73)	(0.72)		(0.55)		(0.54)		(0.58)
ln(mean HH income)	-0.25	-0.13	-0.95	-0.91	-0.37	0.13	-0.30	0.21	-0.25	0.36
	(0.16)	(0.17)	(0.44)	(0.44)	(0.14)	(0.27)	(0.12)	(0.24)	(0.10)	(0.21)
Gini coefficient, HH income	-3.58	-3.11	0.35	0.22	-1.43	-2.20	-1.18	-2.58	-0.70	-2.23
	(0.81)	(0.90)	(0.28)	(0.27)	(0.46)	(1.30)	(0.46)	(1.23)	(0.36)	(0.86)
Pct. Asian	1.88	-0.84	-0.66	-0.65	-0.63	-0.51	-1.59	0.77	-1.55	0.30
	(1.00)	(0.34)	(1.08)	(1.07)	(0.48)	(0.54)	(0.64)	(1.07)	(0.59)	(1.07)
Pct. Black	0.83	0.63	-0.01	0.16	-0.23	0.46	-0.52	1.00	-0.54	0.77
	(0.41)	(0.31)	(0.15)	(0.13)	(0.19)	(0.42)	(0.22)	(0.52)	(0.19)	(0.47)
Pct. Hispanic	0.09	0.29	-3.32	-2.62	-0.27	0.60	-0.31	0.94	-0.42	0.86
	(0.18)	(0.18)	(0.95)	(0.72)	(0.26)	(0.39)	(0.28)	(0.46)	(0.28)	(0.47)
Racial homogeneity index (0=homog., 1=heterog.)	-0.18	0.61	-1.06	-1.09	-3.52	2.83	-5.49	5.33	-5.84	3.36
	(0.49)	(1.47)	(0.50)	(0.50)	(1.18)	(1.97)	(1.24)	(2.52)	(1.28)	(2.73)
Ethnic homogeneity index	0.70	-0.27	0.64	0.60	2.84	-2.08	4.25	-3.82	4.57	-2.43
	(0.70)	(1.12)	(0.35)	(0.33)	(0.90)	(1.49)	(0.93)	(1.86)	(0.97)	(2.00)
Pct. Some college	-1.00	-0.92	0.34	0.35	0.45	-0.94	0.64	-1.37	0.67	-1.03
	(0.41)	(0.38)	(0.21)	(0.21)	(0.40)	(0.60)	(0.34)	(0.63)	(0.33)	(0.64)
Pct. College graduates	0.95	0.67	0.89	0.64	0.68	0.03	0.69	-0.29	0.51	-0.34
	(0.42)	(0.38)	(1.80)	(1.85)	(0.28)	(0.47)	(0.30)	(0.39)	(0.25)	(0.33)
Educational homogeneity index	-2.95	-0.82	-0.54	-0.33	0.67	-1.36	0.90	-2.33	1.26	-2.24
	(1.03)	(0.98)	(1.36)	(1.39)	(0.46)	(1.36)	(0.53)	(1.74)	(0.47)	(1.77)
Individual covariates										
ln(Family income)					0.003		0.000		0.004	
					(0.008)		(0.009)		(0.006)	
Asian					0.003		0.015		0.008	
					(0.020)		(0.019)		(0.023)	
Hispanic					-0.012		-0.019		0.010	
					(0.020)		(0.022)		(0.022)	
Black					-0.012		-0.011		-0.003	
					(0.019)		(0.021)		(0.018)	
Female					-0.011		-0.006		0.007	
					(0.008)		(0.010)		(0.009)	
Parents' highest ed is BA+					-0.015		-0.007		-0.009	
					(0.009)		(0.011)		(0.013)	
Parents' highest ed is some college					-0.012		-0.011		-0.016	
					(0.009)		(0.015)		(0.011)	
N	316	310	304	304	5,475		5,934		6,688	
R2	nr	0.516	0.506	0.517	0.575		0.568		0.583	

Notes: All models include census division fixed effects. Cluster S.E.s are reported for indiv.-level models. "nr"=not reported. Published specification (col. 1) also includes controls for the shares of the MSA population that are 0-19 and 65+ years old.

Appendix Table C3: Full coefficient vectors for alternative first stage models in Table 3

Total stream defn.	**Stream mouths**			**All streams**											
Larger stream defn.	**n/a**			**Hoxby**			**n/a**			**Inter-county**			**>3.5 miles**		
Level	**MSA**		**Indiv**	**MSA**		**Indiv**	**MSA**		**Indiv**	**MSA**		**Indiv**	**MSA**		**Indiv**
	(2A)		**(2B)**	**(3A)**		**(3B)**	**(4A)**		**(4B)**	**(5A)**		**(5B)**	**(6A)**		**(6B)**
Larger streams (100s)				0.037		-0.030				0.260		0.240	0.177		0.190
				(0.021)		(0.019)				(0.055)		(0.047)	(0.036)		(0.029)
Smaller streams (100s)				0.069		0.104				0.014		0.015	0.013		0.001
				(0.013)		(0.013)				(0.016)		(0.013)	(0.017)		(0.013)
Total streams (100s)	0.071		0.064				0.061		0.058						
	(0.013)		(0.011)				(0.010)		(0.009)						
District/MSA covariates	*MSA*	*Dist*	*MSA*	*MSA*	*Dist*	*MSA*	*MSA*	*Dist*	*MSA*	*MSA*	*Dist*	*MSA*	*MSA*	*Dist*	*MSA*
Population (tens of millions)	0.07		0.00	0.07		0.01	0.09		0.02	0.01		-0.04	0.05		-0.01
	(0.16)		(0.17)	(0.16)		(0.16)	(0.15)		(0.16)	(0.15)		(0.18)	(0.15)		(0.16)
Land area (100,000s of sq. mi.)	0.88		0.86	0.82		0.64	0.72		0.71	0.70		0.71	-0.53		-0.59
	(0.72)		(0.72)	(0.73)		(0.54)	(0.71)		(0.67)	(0.69)		(0.60)	(0.79)		(0.52)
ln(mean HH income)	-0.03	-0.27	0.26	0.01	-0.29	0.22	-0.07	-0.30	0.22	-0.03	-0.32	0.28	-0.02	-0.33	0.33
	(0.14)	(0.13)	(0.24)	(0.15)	(0.12)	(0.24)	(0.13)	(0.12)	(0.22)	(0.12)	(0.12)	(0.22)	(0.13)	(0.12)	(0.22)
Gini coefficient, HH income	-3.46	-0.90	-2.80	-3.30	-1.05	-2.54	-3.43	-0.92	-2.89	-3.03	-0.83	-2.53	-2.95	-0.85	-2.22
	(0.89)	(0.48)	(1.32)	(0.93)	(0.47)	(1.20)	(0.84)	(0.47)	(1.19)	(0.83)	(0.48)	(1.18)	(0.84)	(0.46)	(1.17)
Pct. Asian	-1.14	-1.51	0.51	-0.92	-1.54	0.79	-1.01	-1.51	0.62	-0.96	-1.47	0.71	-0.94	-1.55	0.87
	(0.48)	(0.62)	(1.04)	(0.49)	(0.64)	(1.07)	(0.47)	(0.60)	(1.01)	(0.46)	(0.60)	(1.00)	(0.46)	(0.61)	(0.99)
Pct. Black	0.58	-0.49	0.89	0.70	-0.51	1.04	0.61	-0.48	0.93	0.56	-0.45	0.91	0.49	-0.53	0.93
	(0.33)	(0.22)	(0.52)	(0.34)	(0.21)	(0.50)	(0.32)	(0.21)	(0.51)	(0.32)	(0.21)	(0.49)	(0.32)	(0.21)	(0.48)
Pct. Hispanic	0.30	-0.04	0.49	0.37	-0.30	0.92	0.33	-0.07	0.55	0.34	-0.02	0.53	0.34	-0.01	0.55
	(0.20)	(0.30)	(0.45)	(0.20)	(0.27)	(0.44)	(0.19)	(0.28)	(0.41)	(0.19)	(0.27)	(0.39)	(0.19)	(0.25)	(0.38)
Racial homogeneity index	0.27	-4.89	3.43	1.06	-5.29	5.26	0.42	-4.84	3.84	0.24	-4.53	3.72	0.42	-4.87	4.30
	(1.71)	(1.19)	(2.47)	(1.74)	(1.21)	(2.45)	(1.63)	(1.13)	(2.35)	(1.60)	(1.13)	(2.30)	(1.61)	(1.13)	(2.16)
Ethnic homogeneity index	-0.10	3.86	-2.44	-0.69	4.10	-3.77	-0.24	3.82	-2.76	-0.10	3.60	-2.67	-0.23	3.83	-3.04
	(1.30)	(0.89)	(1.82)	(1.31)	(0.91)	(1.80)	(1.23)	(0.84)	(1.72)	(1.21)	(0.84)	(1.70)	(1.21)	(0.84)	(1.60)
Pct. Some college	-1.04	0.68	-1.43	-0.90	0.63	-1.33	-0.93	0.66	-1.40	-0.77	0.66	-1.16	-0.94	0.75	-1.40
	(0.43)	(0.33)	(0.67)	(0.44)	(0.33)	(0.64)	(0.42)	(0.32)	(0.67)	(0.41)	(0.33)	(0.67)	(0.41)	(0.32)	(0.64)
Pct. College graduates	0.34	0.59	-0.22	0.28	0.68	-0.30	0.30	0.60	-0.25	0.23	0.63	-0.35	0.18	0.61	-0.43
	(0.27)	(0.30)	(0.38)	(0.28)	(0.30)	(0.38)	(0.26)	(0.29)	(0.37)	(0.26)	(0.29)	(0.35)	(0.26)	(0.29)	(0.36)
Educational homogeneity index	-0.64	0.81	-2.12	-0.37	0.85	-2.13	-0.62	0.48	-2.13	-0.27	0.48	-1.53	-0.49	0.56	-1.81
	(1.04)	(0.51)	(1.69)	(1.06)	(0.46)	(1.63)	(0.96)	(0.44)	(1.46)	(0.95)	(0.45)	(1.47)	(0.95)	(0.44)	(1.40)
Individual covariates															
ln(Family income)			0.00			0.00			0.00			0.00			0.00
			(0.01)			(0.01)			(0.01)			(0.01)			(0.01)
Asian			0.01			0.02			0.02			0.02			0.01
			(0.02)			(0.02)			(0.02)			(0.02)			(0.02)
Hispanic			-0.01			-0.02			-0.02			-0.01			-0.01
			(0.02)			(0.02)			(0.02)			(0.02)			(0.02)
Black			0.00			-0.01			0.00			0.00			-0.01
			(0.02)			(0.02)			(0.02)			(0.02)			(0.02)
Female			0.00			-0.01			0.00			0.00			0.00
			(0.01)			(0.01)			(0.01)			(0.01)			(0.01)
Parents' highest ed is BA+			-0.01			-0.01			-0.01			-0.01			-0.01
			(0.01)			(0.01)			(0.01)			(0.01)			(0.01)
Parents' highest ed is some college			-0.01			-0.01			-0.01			-0.01			-0.01
			(0.01)			(0.01)			(0.01)			(0.01)			(0.01)
N	319		5,987	311		6,014	335		6,139	335		6,139	335		6,139
R2	0.506		0.531	0.508		0.569	0.505		0.535	0.525		0.557	0.520		0.560

Notes: All models include census division fixed effects. Cluster S.E.s are reported for indiv.-level models, which use the 12th grade reading score samples.

Appendix Table C4: Full coefficient vectors for alternative-instruments IV models in Table 4, 12th grade reading scores

Model	**OLS**		**IV**									
Total stream defn.	**n/a**		**Stream mouths**		**All streams**							
Larger stream defn.	**n/a**		**n/a**		**Hoxby**		**n/a**		**Inter-county**		**>3.5 miles**	
	(1)		**(3)**		**(4)**		**(5)**		**(6)**		**(7)**	
Choice index	-0.25		0.68		4.38		0.87		2.04		1.35	
	(0.79)		(2.79)		(1.98)		(2.59)		(2.36)		(2.30)	
District/MSA covariates	*Dist*	*MSA*	*Dist*	*MSA*	*Dist*	*MSA*	*Dist*	*MSA*	*Dist*	*MSA*	*Dist*	*MSA*
Population (tens of millions)		-0.83		-1.98		-2.65		-1.15		-1.48		-1.28
		(2.01)		(1.42)		(1.36)		(1.40)		(1.37)		(1.38)
Land area (100,000s of sq. mi.)		1.38		0.29		-7.72		-1.18		-3.84		-2.25
		(7.90)		(6.44)		(5.85)		(6.25)		(6.08)		(5.99)
ln(mean HH income)	-0.76	-1.02	0.18	-0.17	1.62	-0.51	-0.41	-1.12	-0.04	-1.21	-0.26	-1.15
	(2.03)	(2.51)	(2.00)	(3.10)	(1.98)	(3.13)	(1.96)	(3.03)	(1.95)	(3.04)	(1.95)	(3.03)
Gini coefficient, HH income	-6.99	-5.20	-2.72	-1.11	3.64	9.07	-5.92	-1.62	-4.82	2.10	-5.48	-0.12
	(8.57)	(15.23)	(8.53)	(21.58)	(8.56)	(19.19)	(8.38)	(20.54)	(8.33)	(19.79)	(8.29)	(19.72)
Pct. Asian	20.31	-32.23	22.77	-34.96	29.64	-38.98	22.23	-33.34	24.22	-34.50	23.03	-33.81
	(11.22)	(15.98)	(6.90)	(10.60)	(6.72)	(10.69)	(6.75)	(10.48)	(6.70)	(10.50)	(6.69)	(10.49)
Pct. Black	13.29	-8.03	13.34	-9.23	16.10	-14.40	13.82	-9.38	14.37	-10.78	14.04	-9.95
	(3.62)	(6.66)	(3.41)	(7.72)	(3.37)	(7.10)	(3.35)	(7.38)	(3.34)	(7.23)	(3.34)	(7.23)
Pct. Hispanic	7.65	-2.31	7.84	-3.72	11.59	-8.70	7.85	-3.05	8.07	-3.82	7.94	-3.36
	(4.17)	(5.42)	(3.47)	(5.20)	(3.53)	(5.01)	(3.41)	(4.98)	(3.41)	(4.92)	(3.40)	(4.91)
Racial homogeneity index	61.54	-16.69	65.47	-25.04	88.45	-49.35	67.11	-22.81	72.90	-29.16	69.45	-25.38
	(19.22)	(28.84)	(18.38)	(29.72)	(17.79)	(28.17)	(17.82)	(29.08)	(17.60)	(28.42)	(17.58)	(28.51)
Ethnic homogeneity index	-44.64	11.16	-47.62	17.02	-64.51	33.82	-49.03	15.42	-53.58	19.85	-50.87	17.21
	(14.35)	(21.30)	(13.99)	(22.26)	(13.49)	(21.35)	(13.53)	(21.75)	(13.35)	(21.33)	(13.33)	(21.39)
Pct. Some college	-4.61	7.60	-4.38	9.35	-4.28	16.36	-5.35	9.91	-6.12	12.29	-5.66	10.87
	(6.92)	(8.96)	(6.94)	(11.27)	(7.00)	(11.12)	(6.85)	(11.09)	(6.86)	(11.03)	(6.85)	(11.00)
Pct. College graduates	7.50	4.42	6.30	4.30	1.59	3.25	6.80	4.26	6.08	4.09	6.51	4.19
	(4.86)	(5.46)	(4.80)	(6.37)	(4.82)	(6.42)	(4.72)	(6.25)	(4.71)	(6.24)	(4.70)	(6.23)
Educational homogeneity index	-13.89	0.07	-7.76	4.83	-17.81	14.05	-14.58	3.66	-15.30	7.38	-14.87	5.16
	(8.93)	(16.05)	(8.32)	(20.97)	(8.18)	(19.70)	(8.09)	(19.65)	(8.07)	(19.57)	(8.05)	(19.58)
Individual covariates												
ln(Family income)	1.75		1.77		1.77		1.75		1.75		1.75	
	(0.17)		(0.17)		(0.17)		(0.17)		(0.17)		(0.17)	
Asian	0.44		0.44		0.34		0.43		0.42		0.43	
	(0.69)		(0.49)		(0.49)		(0.49)		(0.49)		(0.49)	
Hispanic	-2.80		-2.61		-2.51		-2.77		-2.75		-2.76	
	(0.53)		(0.45)		(0.44)		(0.44)		(0.44)		(0.44)	
Black	-4.62		-4.48		-4.60		-4.62		-4.61		-4.62	
	(0.46)		(0.50)		(0.50)		(0.50)		(0.50)		(0.50)	
Female	2.16		2.15		2.18		2.16		2.17		2.17	
	(0.23)		(0.24)		(0.24)		(0.23)		(0.23)		(0.23)	
Parents' highest ed is BA+	5.02		5.05		5.10		5.04		5.06		5.05	
	(0.31)		(0.31)		(0.31)		(0.31)		(0.31)		(0.31)	
Parents' highest ed is some college	2.81		2.82		2.88		2.82		2.84		2.83	
	(0.31)		(0.32)		(0.32)		(0.31)		(0.31)		(0.31)	
N	6,139		5,987		6,014		6,139		6,139		6,139	
R2	0.186		0.186		0.175		0.185		0.184		0.185	

Notes: All models include census division fixed effects. Moulton S.E.s are reported.

Appendix Table C5: Full coefficient vectors for selected models from Table 5

Sample/covariates	Close replication					
Instruments	Hoxby sm. & lg. streams			Inter- & intra-cnty. streams		
Specification	Base samp., no district-lvl covariates	Zip-code matched pub. schls	Pub. & pvt. schools	Base samp., no district-lvl covariates	Zip-code matched pub. schls	Pub. & pvt. schools
	(1)	(2)	(3)	(4)	(5)	(6)
Choice index	4.61	1.40	0.68	1.76	1.10	0.84
	(2.49)	(2.44)	(2.59)	(2.86)	(2.66)	(2.35)
MSA-level covariates						
Population (tens of millions)	-3.55	-0.97	-1.22	-2.02	-1.02	-1.82
	(2.46)	(1.78)	(1.68)	(2.09)	(1.72)	(1.61)
Land area (100,000s of sq. mi.)	-9.44	1.20	0.80	-4.11	0.28	-0.32
	(9.74)	(7.15)	(7.28)	(8.89)	(7.09)	(6.47)
ln(mean HH income)	0.99	-1.50	-0.88	-0.97	-2.22	-0.57
	(2.69)	(2.37)	(2.20)	(2.27)	(2.34)	(2.17)
Gini coefficient, HH income	7.31	-4.28	-0.87	-6.86	-7.00	2.17
	(18.32)	(17.35)	(17.75)	(18.15)	(18.57)	(16.99)
Pct. Asian	-5.35	-5.99	-2.48	-8.13	-7.62	-4.53
	(7.60)	(7.10)	(6.65)	(6.56)	(6.93)	(6.14)
Pct. Black	-1.62	6.20	7.26	0.54	5.31	6.18
	(7.14)	(7.47)	(6.95)	(7.68)	(7.89)	(7.14)
Pct. Hispanic	3.05	7.42	6.84	3.98	7.10	6.35
	(3.67)	(4.09)	(3.81)	(3.71)	(4.20)	(3.87)
Racial homogeneity index	6.46	19.17	14.58	16.62	18.57	4.35
	(33.64)	(29.51)	(28.27)	(29.60)	(28.49)	(27.39)
Ethnic homogeneity index	-5.13	-11.20	-7.08	-12.58	-11.88	0.60
	(25.47)	(22.65)	(21.87)	(21.83)	(21.59)	(20.74)
Pct. Some college	11.91	11.48	7.07	4.87	8.26	6.31
	(10.20)	(9.37)	(9.00)	(9.24)	(9.09)	(8.82)
Pct. College graduates	1.63	9.78	8.39	6.48	12.00	9.80
	(6.66)	(6.12)	(5.90)	(6.13)	(5.83)	(5.46)
Educational homogeneity index	-8.72	-19.84	-17.92	-10.07	-12.45	-14.18
	(21.92)	(19.44)	(18.08)	(20.26)	(18.51)	(17.01)
Individual covariates						
ln(Family income)	1.76	1.07	1.43	1.75	1.12	1.42
	(0.24)	(0.28)	(0.27)	(0.23)	(0.27)	(0.25)
Asian	0.67	0.41	0.42	0.75	0.45	0.34
	(0.68)	(0.79)	(0.74)	(0.65)	(0.77)	(0.71)
Hispanic	-2.37	-2.93	-2.69	-2.68	-3.01	-2.67
	(0.84)	(0.91)	(0.83)	(0.82)	(0.87)	(0.79)
Black	-4.15	-4.85	-5.07	-4.17	-4.88	-5.15
	(0.87)	(0.93)	(0.82)	(0.83)	(0.91)	(0.80)
Female	2.18	2.43	2.63	2.15	2.47	2.59
	(0.39)	(0.43)	(0.40)	(0.37)	(0.41)	(0.38)
Parents' highest ed is BA+	5.48	5.87	6.14	5.39	5.81	6.15
	(0.48)	(0.54)	(0.50)	(0.47)	(0.53)	(0.48)
Parents' highest ed is some college	3.10	3.05	3.08	3.01	3.16	3.21
	(0.52)	(0.57)	(0.53)	(0.49)	(0.55)	(0.51)
N	5,939	5,445	6,670	6,144	5,631	6,900
R2	0.164	0.170	0.191	0.178	0.174	0.199

Notes: All models include census division fixed effects. Clustered S.E.s are reported.

Appendix Table D1: First-stage estimates for alternative instruments, using "preferred" replication sample and covariates

	(1)	(2)	(3)	(4)	(5)	(6)
Total stream definition	**Stream mouths**		**All streams**			
Larger stream definition	**Hoxby**	**n/a**	**Hoxby**	**n/a**	**Inter-county**	**>3.5 miles**
MSA level						
Larger streams (100s)	0.043		0.040		0.250	0.178
	(0.021)		(0.021)		(0.054)	(0.035)
Smaller streams (100s)	0.091		0.068		0.017	0.012
	(0.018)		(0.012)		(0.016)	(0.017)
Total streams (100s)		0.071		0.061		
		(0.013)		(0.010)		
F statistic, instruments	16.3	31.7	17.8	37.9	26.1	25.0
Individual level (12th grade reading sample)						
Larger streams (100s)	0.015		0.001		0.236	0.171
	(0.020)		(0.018)		(0.044)	(0.029)
Smaller streams (100s)	0.114		0.099		0.025	0.018
	(0.018)		(0.012)		(0.014)	(0.017)
Total streams (100s)		0.072		0.066		
		(0.011)		(0.009)		
F statistic, instruments	28.4	44.2	38.8	53.3	34.3	36.7
Individual level (8th grade reading sample)						
Larger streams (100s)	-0.012		-0.017		0.227	0.151
	(0.018)		(0.017)		(0.044)	(0.029)
Smaller streams (100s)	0.132		0.102		0.021	0.018
	(0.017)		(0.012)		(0.014)	(0.017)
Total streams (100s)		0.067		0.060		
		(0.012)		(0.009)		
F statistic, instruments	32.1	33.2	34.7	40.7	28.2	28.1

Notes: Base samples are those from Column 4 of Tables 1 (individual level) and 2 (Panel B; MSA level), though some observations that were excluded from those samples for missing data on larger streams are included here in Columns 2, 4, 5, and 6. In individual-level specifications, standard errors are clustered at the MSA level.

Appendix Table D2: IV estimates of choice effect, using alternative instruments and "preferred" replication sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	**IV**					
Total stream definition	**n/a**	**Stream mouths**		**All streams**			
Larger stream definition	**n/a**	**Hoxby**	**n/a**	**Hoxby**	**n/a**	**Inter-county**	**>3.5 miles**
Panel : 12th grade reading scores							
Choice effect	-0.17	3.29	2.79	4.05	3.41	3.36	2.64
S.E. (Moulton)	(0.62)	(1.83)	(2.44)	**(1.79)**	(2.13)	(1.96)	(1.92)
S.E. (Cluster)	(0.97)	(2.56)	(2.69)	**(2.06)**	(2.45)	(2.60)	(1.98)
p-value, exog. test	--	0.20	0.25	0.04	0.10	0.12	0.10
Panel A: 8th grade reading scores							
Choice index	-0.55	2.93	0.58	2.79	1.03	0.95	0.51
S.E. (Moulton)	(0.62)	(1.58)	(1.95)	(1.55)	(1.83)	(1.72)	(1.65)
S.E. (Cluster)	(0.61)	**(1.40)**	(1.99)	**(1.29)**	(1.82)	(1.39)	(1.53)
p-value, exog. test	--	0.00	0.55	0.00	0.34	0.20	0.44

Notes: Base samples are those from Column 4 of Table 1, though some observations that were excluded from that sample for missing data on larger streams are included here in Columns 3 and 5-7. Exogeneity tests are based on clustered specification. Bold S.E.s indicate that with that S.E., the coefficient is significant at the 5% level.

Appendix Table D3. Exploration of potential bias from exclusion of private school students, 8th grade reading scores

Covariate specification	Close replication			Preferred replication		
Streams instruments	OLS	Hoxby	Inter- and intra-cnty	OLS	Hoxby	Inter- and intra-cnty
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Base specification, replication sample of public school students						
Choice index coefficient	-0.06	**5.93**	1.67	-0.55	**2.93**	0.95
	(0.82)	(2.32)	(1.77)	(0.61)	(1.40)	(1.39)
N	10,709	10,429	10,709	12,049	11,719	12,049
p-value, exog. test		0.00	0.21		0.00	0.20
Panel B: Without district-level covariates						
Choice index coefficient	-0.13	**5.03**	1.98	-0.60	2.06	0.97
	(0.80)	(2.33)	(1.82)	(0.60)	(1.28)	(1.40)
N	10,729	10,449	10,729	12,049	11,719	12,049
p-value, exog. test		0.01	0.15		0.03	0.19
Panel C: Public school students in zip-code matched sample (no district covariates)						
Choice index coefficient	-0.87	1.93	0.19	-0.59	1.57	0.20
	(0.70)	(1.55)	(1.50)	(0.60)	(1.30)	(1.35)
N	10,394	10,117	10,394	11,992	11,662	11,992
p-value, exog. test		0.04	0.39		0.08	0.50
Panel D: Public and private school students in zip code-matched sample						
Choice index coefficient	-0.37	1.07	0.65	-0.13	0.78	0.45
	(0.65)	(1.56)	(1.57)	(0.59)	(1.40)	(1.48)
N	13,879	13,482	13,879	16,026	15,558	16,026
p-value, exog. test		0.30	0.44		0.48	0.66

Notes: Clustered standard errors and test statistics are reported. Bold coefficients are significant at the 5% level.

Appendix Table D4: Choice effect estimates for all six NELS test scores

Sample	Hoxby/NELS		Close replication			Preferred replication		
Model		IV		IV			IV	
Instruments	OLS	Hoxby streams	OLS	Hoxby streams	Inter- and intra-county streams	OLS	Hoxby streams	Inter- and intra-county streams
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: 8th grade								
Reading	-0.65	**4.45**	-0.06	**5.93**	1.67	-0.55	**2.93**	0.95
	(0.71)	(1.99)	(0.82)	(2.32)	(1.77)	(0.61)	(1.40)	(1.39)
Mathematics	-0.61	**4.23**	-0.55	**4.52**	-0.64	-0.80	**3.22**	-0.60
	(0.73)	(1.79)	(0.71)	(1.77)	(1.57)	(0.63)	(1.59)	(1.41)
Panel B: 10th grade								
Reading	-0.85	**6.73**	-0.88	**6.19**	5.34	0.19	3.94	4.39
	(1.01)	(2.51)	(1.06)	(2.64)	(2.90)	(0.95)	(2.43)	(2.43)
Mathematics	-0.89	**7.91**	-1.10	**4.95**	1.32	0.10	2.67	0.00
	(0.93)	(2.21)	(0.86)	(2.14)	(2.32)	(0.75)	(1.99)	(2.08)
Panel C: 12th grade								
Reading	-1.24	5.30	-0.25	**4.74**	2.04	-0.17	3.29	3.36
	(1.15)	(2.94)	(0.94)	(2.42)	(2.94)	(0.97)	(2.56)	(2.60)
Mathematics	-1.21	3.49	-0.34	2.48	1.56	0.60	2.15	2.69
	(0.93)	(2.29)	(0.80)	(2.08)	(2.29)	(1.01)	(2.06)	(2.30)

Notes: Cluster S.E.s in parentheses. Bold coefficients are significant at the 5% level.

Appendix Table D5: MSA-level first stage estimates, using only MSAs in the NELS 12th grade sample

	(1)	(2)	(3)	(4)	(5)	(6)
Total stream definition	**Stream mouths**		**All streams**			
Larger stream definition	**Hoxby**	**n/a**	**Hoxby**	**n/a**	**Inter-county**	**>3.5 miles**
Panel A: Hoxby/NELS sample						
Larger streams (100s)	-0.044					
	(0.028)					
Smaller streams (100s)	0.143					
	(0.025)					
Total streams (100s)		0.057				
		(0.016)				
F statistic (instruments)	16.0	13.0				
Panel B: Close replication sample						
Larger streams (100s)	-0.013		-0.018		0.242	0.166
	(0.028)		(0.028)		(0.064)	(0.043)
Smaller streams (100s)	0.143		0.113		0.024	0.021
	(0.024)		(0.018)		(0.020)	(0.023)
Total streams (100s)		0.073		0.066		
		(0.016)		(0.013)		
F statistic (instruments)	18.3	21.7	20.2	26.8	17.8	16.4
Panel C: Preferred replication sample						
Larger streams (100s)	0.019		0.005		0.245	0.164
	(0.026)		(0.026)		(0.062)	(0.042)
Smaller streams (100s)	0.121		0.106		0.029	0.029
	(0.022)		(0.017)		(0.019)	(0.022)
Total streams (100s)		0.078		0.071		
		(0.015)		(0.012)		
F statistic (instruments)	16.9	26.6	20.8	33.4	21.4	19.3

Appendix Table D6: Two-sample IV estimates of choice effects on 12th grade reading scores

Sample	Hoxby/NELS	Close replication		Preferred replication	
Streams instruments	Hoxby	Hoxby	Inter- and intra-cnty	Hoxby	Inter- and intra-cnty
	(1)	(2)	(3)	(4)	(5)
Panel A: 8th grade reading scores					
First stage uses only NELS MSAs					
Choice index coefficient	**3.99**	**4.22**	0.76	**2.70**	0.90
	(1.60)	(1.75)	(1.51)	(1.23)	(1.30)
First stage uses all MSAs					
Choice index coefficient	3.68	3.46	0.80	2.15	0.90
	(2.27)	(2.15)	(1.51)	(1.68)	(1.28)
Panel B: 12th grade reading scores					
First stage uses only NELS MSAs					
Choice index coefficient	4.58	**4.64**	2.31	3.09	3.15
	(2.61)	(2.20)	(2.24)	(2.51)	(2.36)
First stage uses all MSAs					
Choice index coefficient	2.14	4.15	2.71	3.69	3.41
	(2.98)	(2.57)	(2.27)	(2.75)	(2.64)

Notes: First stage is estimated at MSA level, on the indicated sample of MSAs. Reported coefficients are those on the first stage fitted value from the second stage regression. Reported standard errors are clustered S.E.s from the second stage regression. These are unadjusted for the presence of a generated regressor, so are not correct, and are likely downward-biased estimates of the true standard errors. Bold coefficients are significant at the 5% level with these S.E.s.

Appendix Table D7: Reduced-form estimates, effect of streams on test scores

	8th grade reading scores						12th grade reading scores					
Total stream definition	**Stream mouths**		**All streams**				**Stream mouths**		**All streams**			
Larger stream definition	**Hoxby**	**n/a**	**Hoxby**	**n/a**	**Inter-county**	**>3.5 miles**	**Hoxby**	**n/a**	**Hoxby**	**n/a**	**Inter-county**	**>3.5 miles**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**
Panel A: Hoxby/NELS sample												
Larger streams (100s)	-0.50						**-1.39**					
	(0.26)						(0.55)					
Smaller streams (100s)	**0.60**						**0.77**					
	(0.23)						(0.38)					
Total streams (100s)		0.09						-0.18				
		(0.16)						(0.23)				
Panel B: Close replication sample												
Larger streams (100s)	-0.36		-0.37		0.50	-0.21	**-1.00**		**-1.02**		0.57	0.52
	(0.20)		(0.20)		(0.41)	(0.28)	(0.41)		(0.40)		(1.18)	(0.49)
Smaller streams (100s)	**0.36**		**0.26**		-0.10	0.13	**0.58**		**0.44**		-0.25	-0.42
	(0.16)		(0.12)		(0.15)	(0.18)	(0.28)		(0.20)		(0.28)	(0.32)
Total streams (100s)		0.00		0.01				-0.17		-0.09		
		(0.11)		(0.09)				(0.16)		(0.14)		
Panel C: Preferred replication sample												
Larger streams (100s)	-0.30		-0.30		0.53	-0.15	-0.54		-0.55		0.97	0.47
	(0.19)		(0.18)		(0.38)	(0.27)	(0.34)		(0.33)		(1.10)	(0.53)
Smaller streams (100s)	**0.34**		**0.25**		-0.08	0.14	**0.60**		**0.52**		-0.08	-0.07
	(0.14)		(0.10)		(0.13)	(0.16)	(0.27)		(0.19)		(0.24)	(0.32)
Total streams (100s)		0.02		0.03				0.06		0.12		
		(0.10)		(0.08)				(0.15)		(0.14)		

Note: Regressions are estimated by OLS, at individual level, with clustered standard errors. Bold coefficients are significantly different from zero at the 5% level.

Appendix Table D8: MSA-level estimates of the choice effect

	12th grade reading scores							8th grade reading scores						
	OLS	**IV**						**OLS**	**IV**					
Total strean	**n/a**	**Mouths**		**All streams**				**n/a**	**Mouths**		**All streams**			
Larger strea	**n/a**	**Hoxby**	**n/a**	**Hoxby**	**n/a**	**Inter-county**	**>3.5 miles**	**n/a**	**Hoxby**	**n/a**	**Hoxby**	**n/a**	**Inter-county**	**>3.5 miles**
CD Sample														
All controls	-0.48	4.30	-3.77	4.26	-2.60	-0.92	-0.74	-0.79	3.57	0.42	3.42	0.87	-0.41	-2.08
	(1.20)	(3.15)	(5.13)	(3.39)	(4.82)	(3.71)	(3.88)	(0.75)	(2.18)	(3.54)	(2.31)	(3.28)	(2.28)	(2.52)
MSA controls	-1.33	3.89	-2.56	3.83	-1.84	0.37	0.09	-0.56	3.93	1.43	4.29	2.12	3.37	2.95
	(1.29)	(3.25)	(4.62)	(3.46)	(4.45)	(3.60)	(3.82)	(0.99)	(2.62)	(3.74)	(2.82)	(3.60)	(2.94)	(3.09)
Close replication sample														
All controls	0.45	5.08	0.57	4.72	0.94	2.21	1.81	-0.17	**5.50**	1.98	**4.98**	2.53	1.10	-0.43
	(1.14)	(3.03)	(3.79)	(2.87)	(3.45)	(2.96)	(2.86)	(0.86)	(2.60)	(3.48)	(2.43)	(3.14)	(2.39)	(2.36)
MSA controls	-1.26	3.74	1.50	2.58	0.88	2.39	1.64	0.04	5.23	3.62	3.69	2.90	3.22	3.20
	(1.26)	(3.30)	(4.28)	(3.11)	(3.86)	(3.32)	(3.29)	(1.07)	(2.75)	(3.68)	(2.58)	(3.29)	(2.77)	(2.80)
Preferred replication sample														
All controls	0.17	3.38	3.19	4.39	3.89	3.24	3.22	-0.38	2.68	0.46	3.01	1.21	1.04	-0.04
	(1.19)	(3.38)	(3.82)	(3.08)	(3.42)	(2.94)	(2.92)	(0.74)	(2.06)	(2.71)	(1.98)	(2.47)	(2.01)	(2.10)
MSA controls	-0.81	2.73	3.53	2.49	3.20	3.24	2.66	-0.57	2.75	2.05	1.82	1.74	2.54	2.47
	(1.32)	(3.42)	(3.89)	(3.13)	(3.52)	(3.17)	(3.28)	(1.00)	(2.42)	(3.05)	(2.33)	(2.81)	(2.53)	(2.63)

Notes: "All controls" specifications include MSA-level averages of individual- and district-level covariates, computed within the NELS sample. "MSA controls" specifications exclude the individual- and district-level covariates. Bold coefficients are significant at the 5% level.

Appendix Table E1: Estimates of choice effect on ln(per pupil spending)

	District-level analysis			MSA-level analysis		
	OLS	IV, Hoxby streams	IV, Inter- and intra-county streams	OLS	IV, Hoxby streams	IV, Inter- and intra-cnty streams
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Published estimates						
Choice index coefficient	-0.072	-0.076				
S.E. (Moulton)	(0.022)	(0.034)				
Panel B: Hoxby/NCES data						
Choice index coefficient	-0.070	-0.074				
S.E. (Classical)	(0.012)	(0.037)				
S.E. (Moulton)	(0.024)	(0.141)				
S.E. (Cluster)	(0.037)	(0.111)				
Panel C: Replication, "close" covariates						
Choice index coefficient	-0.121	-0.201	-0.139	-0.124	-0.187	-0.135
S.E. (Classical)				(0.035)	(0.111)	(0.092)
S.E. (Moulton)	(0.023)	(0.154)	(0.084)			
S.E. (Cluster)	(0.034)	(0.111)	(0.137)			
Panel D: Replication, "preferred" covariates						
Choice index coefficient	-0.120	-0.211	-0.142	-0.119	-0.194	-0.120
S.E. (Classical)				(0.035)	(0.111)	(0.092)
S.E. (Moulton)	(0.023)	(0.153)	(0.084)			
S.E. (Cluster)	(0.034)	(0.112)	(0.131)			

Notes: Dependent variable is log of average per pupil spending in the district (columns 1-3, N=5,336-5,804) or MSA (columns 4-6, N=302-333). "Hoxby streams" instruments are Hoxby's larger and smaller streams variables. All specifications include usual list of MSA-level covariates and division fixed effects; those in columns 1-3 also include usual district-level covariates.