Online Appendix for
"The Welfare Effects of Coordinated Assignment:
Evidence from the New York City High School Match"

Atila Abdulkadiroğlu Nikhil Agarwal Parag A. Pathak

November 2016

A Appendix: Computational (Not for Publication)

The demand model is an ordered version of the model in Rossi, McCulloch, and Allenby (1996). We assume that the utility for student i for program j can be written as:

$$\begin{aligned} u_{ij} &= \delta_j + \sum_l \alpha^l z_i^l x_j^l + \sum_k \gamma_i^k x_j^k - d_{ij} + \varepsilon_{ij}, \\ \text{with} \quad \delta_j &= x_j\beta + \xi_j. \end{aligned}$$

We parametrize the random coefficients as follows:

$$\gamma_i \sim \mathcal{N}(0, \Sigma_\gamma), \qquad \xi_j \sim \mathcal{N}(0, \sigma_\xi^2), \qquad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2).$$

The priors for β, α, Σ_γ, σ_ξ^2, and σ_ε^2 are as follows:

$$\beta \sim \mathcal{N}(0, \bar{\Sigma}_\beta), \qquad \alpha \sim \mathcal{N}(0, \bar{\Sigma}_\alpha)$$

$$\Sigma_\gamma \sim \text{IW}(\bar{\Sigma}_\gamma, \nu_\gamma), \quad \sigma_\xi^2 \sim \text{IW}(\bar{\sigma}_\xi^2, \nu_\xi), \quad \text{and} \quad \sigma_\varepsilon^2 \sim \text{IW}(\bar{\sigma}_\varepsilon^2, \nu_\varepsilon),$$

where IW is the inverse Wishart distribution. Following Chapter 5 of Rossi, Allenby, and McCulloch (2005), we set diffuse priors as follows: the prior variances of β and α are 100 times the identity matrix, and

$$(\bar{\Sigma}_\gamma, \nu_\gamma) = ((3 + \dim(\gamma_i))I_{\dim(\gamma_i)}, 3 + \dim(\gamma_i)),$$

$$(\bar{\sigma}_\xi^2, \nu_\xi) = (1, 2) \quad \text{and} \quad (\bar{\sigma}_\varepsilon^2, \nu_\varepsilon) = (3 + J, 3 + J),$$

where I_k is the identity matrix of dimension k.

The Gibbs sampler iterates through the following steps (where, for notational simplicity, we omit conditioning on the observed data and the priors). First, we iterate through the observed rank-ordered lists to update the values of u_{ij}. We then draw utilities for the unranked options by observing that their indirect utility must be at most the indirect utility of the lowest ranked option. This step can be written as

$$u_{ij} | u_{i-j}, \mathbf{r}_i, \beta, \xi, \gamma_i, \alpha,$$

where each simulation is from a (two-sided) truncated normal.

Given the utilities, the posteriors of ξ, β and α are multivariate normal distributions that can be computed as follows:

$$\begin{aligned} \xi &\mid u, \gamma, \beta, \alpha, \sigma_\xi^2, \\ \beta &\mid u, \gamma, \xi, \alpha, \bar{\Sigma}_\beta, \\ \alpha &\mid u, \gamma, \beta, \alpha, \bar{\Sigma}_\alpha, \end{aligned}$$

where u and γ stack the utilities and random coefficients for all students. We then update the student-specific random coefficients:

$$\gamma_i | u_i, \beta, \xi, \alpha, \Sigma_\gamma.$$

The priors and distribution of ε_{ij} imply that a posterior is a multivariate normal distribution for each student. Finally, we sample from the posteriors $\sigma_\varepsilon^2|\varepsilon$, $\sigma_\xi^2|\xi$ and $\Sigma_\gamma|\gamma$, which are given by inverse Wishart distributions.

For the Full sample estimates in the main specification, we iterate through the Markov Chain 1.25 million times, and discard the first 0.75 million draws as "burn in" to ensure mixing. We diagnosed mixing by examining trace plots and computing the Potential Scale Reduction Factor (PSRF) following Gelman and Rubin (1992). Because of computational constraints in drawing from separate chains, we split the draws after the burn-in period into three equally sized continguous pieces and computed the PSRF using the first and third pieces. The PSRFs for almost all parameters were within 1.1 and were within 1.3 for all parameters. Trace plots for the few parameters with PSRFs higher than 1.1 did not indicate any obvious convergence issues.

Estimates of the 10% samples were computed by iterating through the Markov Chain 1 million times and discarding the first 0.75 million draws. We obtained estimates from three distinct chains initiated from dispersed starting values. We compared variances within each chain and the variance between chains, by computing both within and across PSRF chain values. For nearly all parameters, the PSRF is close to one, which suggests we have reached the target distribution.

Our estimates report the posterior mean and standard deviations. We examined the histograms of the marginal distributions of the posteriors to assess the skew. These histograms indicate that the means, modes and medians of the parameters in the main specification are similar.

B Appendix: Subway Distances (Not for Publication)

In New York, high school students who live within 0.5 miles of a school are not eligible for transportation. If a student lives between 0.5 and 1.5 miles from a school, the Metropolitan Transit Authority provides them with a half-fare student Metrocard that works only for bus transportation. If they reside 1.5 miles or more from a school, they obtain full-fare transportation with a student Metrocard that works for subways and buses and is issued by the school transportation office.

Since subway is a common mode of transportation in New York City, this appendix assesses how the driving distance measure we utilize in the paper differs from commuting distance using NYC's subway system. Subway distance is defined as the sum of distance on foot to the student's nearest subway station, travel distance on the subway network to a school's nearest subway station, and the distance on foot from that station to the school. To compute these distances, we used ESRI's ArcGIS software and information on the NYC subway system from GIS files downloaded from Metropolitan Transit Authority's website. Details on these sources are in the Data appendix.

The overall correlation between driving distance and total commuting distance for all student-program pairs is 0.96. A regression of commuting distance on driving distance yields a coefficient of 0.77. Table B2 provides a summary of the correlations by the student and school borough. The correlations are higher than 0.84, except for schools in Staten Island, where the subway system is not quite as extensive as in other boroughs. In fact, it may be that driving distances are a more accurate measure than subway distance of Staten Island.

Panels B and C show that most students are assigned to schools in their borough in both the uncoordinated and coordinated mechanisms. In both mechanisms, a very small number of students who do not live in Staten Island are assigned to schools there, and conversely, only a small number of students living in Staten Island are assigned to schools in a different borough.

Table B1. Offer Processing in the Second Year of the Coordinated Mechanism (2004-05)

		Distance to School (in miles)			
	Number of Students	Assignment	Enrollment	Exit from NYC Public Schools	In NYC Public, but at School Other than Assigned
	(1)	(2)	(3)	(4)	(5)
			Coordinated Mechanism - 2004 - 2005		
Overall	69,013	4.07	3.96	6.6%	6.9%
Main Round	60,251	4.11	3.99	6.5%	6.4%
Supplementary Round	5,475	4.16	4.03	8.5%	13.6%
Administrative Round	3,287	3.25	3.26	4.9%	5.4%

Notes: Columns 2-5 report means. Coordinated mechanism for 2004-05 based on deferred acceptance. Student distance is calculated as road distance using ArcGIS. Assignment is the school assigned at the conclusion of the high school assignment process. Enrollment is the school in which a student enrolls in October following application. Assigned students exit New York City if they are not enrolled in any NYC public high school in October following application. Enrolled in School other than Assigned means the student is in NYC Public but in a school other than that assigned at end of match. Final assignment round is the round during which an offer to the final assigned school is first made.

Table B2. Subway and Driving Distance and Cross-Borough Travel

	School Borough					
	Bronx	Brooklyn	Manhattan	Queens	Staten Island	Total
Student Borough	(1)	(2)	(3)	(4)	(5)	(6)
	A. Correlation between Subway and Driving Distance					
Bronx	0.90	0.93	0.97	0.91	0.76	…
Brooklyn	0.90	0.91	0.95	0.91	0.92	…
Manhattan	0.96	0.95	0.98	0.95	0.76	…
Queens	0.91	0.91	0.95	0.87	0.85	…
Staten Island	0.84	0.92	0.85	0.89	0.73	…
	B. Cross-Borough Travel in Uncoordinated Mechanism					
Bronx	15,187	41	1,382	66	1	16,677
Brooklyn	13	20,877	1,073	502	12	22,477
Manhattan	89	42	8,604	24	1	8,760
Queens	15	493	586	16,498	0	17,592
Staten Island	2	13	59	4	4,774	4,852
	C. Cross-Borough Travel in Coordinated Mechanism					
Bronx	13,335	85	2,049	84	8	15,561
Brooklyn	39	20,035	1,858	846	40	22,818
Manhattan	238	108	7,492	52	7	7,897
Queens	26	584	1,028	14,972	9	16,619
Staten Island	3	37	69	4	3,913	4,026

Notes: Panel A reports on the correlation between student-school distance as computed by road distance and subway distance. Subway distance is the sum of distance on foot to the student's nearest subway station, travel distance on the subway network to a school's nearest subway station, and the distance on foot from that station to the school. Both distance measures are computed using ArcGIS. Panels B and C report on the number of students in each borough who are assigned to schools in each borough.

Table B3. Main Round Assignments in Coordinated Mechanism, by Length of Rank Order List

		Length of Rank Order List											
Choice Assigned	All	1	2	3	4	5	6	7	8	9	10	11	12
Total	69,907	4,597	3,282	4,128	4,622	4,952	4,776	4,406	4,390	4,558	6,135	9,849	14,212
1	31.9%	88.6%	40.7%	35.2%	31.9%	27.9%	28.6%	27.1%	25.7%	25.6%	25.4%	26.2%	25.2%
2	15.0%		39.8%	17.7%	15.1%	14.8%	14.6%	13.7%	13.9%	13.9%	15.2%	14.7%	14.6%
3	10.2%			24.3%	11.6%	11.6%	10.6%	10.0%	10.8%	9.9%	10.4%	10.4%	10.5%
4	7.3%				18.0%	9.3%	8.1%	7.9%	8.0%	7.6%	7.6%	7.8%	8.2%
5	5.4%					12.8%	7.0%	7.0%	6.3%	6.1%	6.6%	6.2%	6.7%
6	3.9%						10.2%	5.7%	4.9%	5.0%	4.9%	4.8%	5.3%
7	2.9%							8.1%	4.3%	4.4%	4.0%	4.1%	4.3%
8	2.0%								5.8%	3.4%	3.3%	2.9%	3.5%
9	1.5%									4.0%	2.8%	2.7%	2.8%
10	1.1%										3.2%	2.3%	2.6%
11	0.8%											2.6%	2.2%
12	0.5%												2.5%
Unassigned	17.5%	11.4%	19.5%	22.8%	23.3%	23.6%	20.9%	20.6%	20.3%	20.1%	16.7%	15.3%	11.6%

Notes: This table reports choices assigned after the main round in the coordinated mechanism in 2003-04.

Table B4. Assignment and Enrollment Decisions of Students in Coordinated Mechanism by Rank Order List Length

		Length of Rank Order List											
	All	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th
						A. Students Offered Assignment in Main Round							
Number of Students	57,658	4,072	2,641	3,187	3,545	3,782	3,776	3,497	3,499	3,642	5,113	8,340	12,564
Average Rank of Assignment	3.00	1.00	1.49	1.86	2.21	2.53	2.76	3.04	3.20	3.35	3.49	3.60	3.93
Accept Main Round Assignment	92.7%	91.2%	88.5%	88.4%	90.2%	91.2%	92.3%	91.9%	93.0%	93.6%	94.5%	94.6%	94.3%
Enroll in Private School	2.5%	6.9%	7.4%	6.1%	4.5%	2.9%	2.4%	2.1%	1.9%	1.2%	1.0%	0.7%	1.0%
Remain in Current School	1.2%	1.2%	2.0%	2.3%	1.9%	2.1%	1.4%	1.8%	1.4%	1.2%	0.9%	0.7%	0.6%
Attend Specialized or Alternative School	0.1%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.2%	0.1%	0.1%	0.2%	0.1%	0.0%
Participate in Supplementary Round	0.3%	0.1%	0.2%	0.2%	0.4%	0.5%	0.6%	0.3%	0.6%	0.3%	0.3%	0.2%	0.3%
						B. Students Unassigned after Main Round							
Number of Students	12,249	525	641	941	1,077	1,170	1,000	909	891	916	1,022	1,509	1,648
Participate in Supplementary Round	52.6%	26.1%	44.8%	54.0%	54.1%	56.2%	55.6%	55.7%	52.7%	46.5%	43.5%	49.6%	68.2%
Enroll at Supplementary Round Assignment	72.9%	73.0%	85.0%	76.0%	75.5%	77.8%	73.0%	75.9%	74.5%	68.8%	71.7%	69.5%	66.3%
Enroll in Private School	2.8%	6.7%	6.1%	4.7%	3.5%	3.8%	2.2%	1.7%	1.6%	1.9%	2.2%	1.4%	1.9%
Remain in Current School	3.2%	6.7%	6.2%	5.6%	5.3%	4.4%	3.2%	3.3%	2.5%	2.0%	1.5%	0.9%	1.8%
Attend Specialized or Alternative School	0.3%	0.8%	0.5%	0.6%	0.2%	0.1%	0.3%	0.3%	0.3%	0.1%	0.2%	0.5%	0.1%

Notes: Assignment and enrollment decisions of students in the demand estimation sample under the coordinated mechanism. Panel A restricts to students who received an assignment to an NYC Public School in the Main Round. Panel B restricts to students who did not receive an assignment in the Main Round.

C Appendix: Data (Not for Publication)

The data for this study come from the NYC Department of Education (DOE), the 2000 US Census, ArcGIS Business Analyst toolbox, and GFTS NYC subway data from the NYC Metropolitan Transit Authority. These sources provide us with data on students, schools, the rank-order lists submitted by students, assignments of students to schools, or the distance between students and schools on either the road network or the subway system. Students and programs are uniquely identified by a number that can be used to populate fields and merge across DOE datasets. We geocode student and school addresses to merge with geo-spatial data.

We use three samples of students in our analysis: one sample to estimate demand and two to infer the welfare effects of the mechanism change. The welfare samples consist of public middle school students who matriculate into NYC Public High Schools in the academic years 2003-04 and 2004-05. The demand sample consists of public middle school students who participated in the Main round of the mechanism in 2003-04. The demand sample and the welfare sample from 2003-04 are not nested because students participating in the mechanism may choose to enroll in schools outside the NYC Public School system, whereas other students may be assigned to public schools outside the main assignment process.

C.1 Students

Assignment and Rank Data

Data on the assignment system come from the DOE's enrollment office. The files indicate all final assignments of students in both analysis years. We use these assignments as the basis of our baseline welfare calculations. In addition, the assignment system also provides separate files that detail the rank orders, applications, or processes through which a student is assigned to a given school.

We use the records from the Main round in the new mechanism to obtain the rank-order lists submitted by students and the assignment proposed by the mechanism. A total of 87,355 students participated in the main round.

For the old mechanism, the assignment system provides student choice and decision files for the Main round. The former contains the ranked applications submitted by the students and the latter provides the school decisions to accept/reject/waitlist students and the students' responses to these offers, if any. A total of 84,272 students participated in the Main round.

The old assignment system also contains several files documenting the supplementary variable assignment process (VAS) round.

Assignment Rounds and Offers in the Old Mechanism

The files in the old mechanism do not contain direct information on how students were assigned to their programs. However, we are able to determine whether a student applied to a particular program/school in the Main process or the supplementary VAS process. We first append fields indicating whether a student applied to her assigned program in the main process. We also append a field indicating whether a student applied to her assigned school in the supplementary

VAS process. It turns out that no final assignment appears in both the Main and the VAS files. We therefore categorize the former assignments as Main-round assignments and the latter as VAS assignments. We assume all other assignments occur in the Administrative round. Based on conversations with DOE officials, we surmise that students were typically assigned to the school closest to home that had open seats. Our understanding is that most students who participated in the VAS process did not have a default local school. Analyzing the geographic distribution if students assigned administratively, as per our definition, supports this; many parts of NYC have no students assigned administratively.

Finally, we also append the number of offers made to a particular student using a file with the initial school response to the student application.

Assignment Rounds in the New Mechanism

We use the NYC assignment files described above to determine the process through which a student was assigned a given school.

The assignment files in the new mechanism contain, for every student program-pair ranked in either the Main or the Supplementary round, two fields indicating whether the student is eligible for the school and if the student was assigned to that school. A final assignment is treated as a Main round assignment if it appears as an eligible assignment in the Main round. Assignments not made in the Main round are treated as Supplementary round assignments if they appear in the Supplementary round files. All other assignments are treated as Administrative assignments.

Student Characteristics

The records from the NYC Department of Education contain students' street address, previous and current grade, gender, ethnicity, and whether the student was enrolled in a public middle school. Each student is identified by a unique number that allows us to merge these data with additional NYC DOE data on student scores in middle school standardized tests, Limited English Proficiency status, and Special Education status. A separate file indicates subsidized lunch status as of the 2004-05 enrollment. If a student is not in that file, we code the student as not receiving a subsidized lunch.

There are several standardized tests taken by middle school students in NYC. To avoid the concern that two different tests may not be comparable indicators of student achievement, we identify the modal standardized math and reading tests taken by students in our sample. These are the May tests with codes CTB and TEM respectively. Of the students who did not take either of these tests in May, at most 10% (<2% of the full sample) took a different standardized test in the same subject while in middle school. We verify that test score distribution and support are similar across the two years in our sample. Some students took the test multiple times. The highest score obtained by a student was used in these instances.

In 2002-03, the math and reading scores are missing for 13.56% and 17.55% of students, respectively, from our final sample. For the 2003-04 welfare sample, scores are missing for 8.29% and 13.57% students, respectively, for math and reading. In the demand sample the corresponding fractions are 7.13% and 12.56%.

Geographic Data

We use the 2000 US Census to obtain block group family income. Student home address and distances to school were calculated using ArcGIS. Corrections to the addresses, when necessary, were made using Google Map Tools followed by manual checks and corrections.

The final address set was geocoded using ArcGIS geocoder with the address-set in the Business Analyst toolbox (ver. 10.0). We first used an exact match to determine if a student's address can be geocoded precisely to a rooftop. If the results were unreliable, we coded the student to the zip code centroid. The vast majority of students were rooftop geocoded. The OD Cost matrix tool in the Network Analyst toolbox was used to compute the distance by road for each student-school pair. The road network is also obtained from Business Analyst.

Our computation of subway distances assumes that a student first walks to the closest subway stop, then uses the subway system to travel to the subway stop closest to the school, and finally walks from the subway to the school. The Subway stop locations are taken from the GTFS and geodata data on the NYC Metropolitan Transit Authority website. The Network Analyst toolbox is used to compute the walking distance and the GTFS data is used to compute the distance on the subway system between every pair of subway stops.

Merging Student Records

Assignment and other DOE files are matched using the unique student identifier linking these data. Each 8th-grade non-private middle school student in the Department of Education records could be merged uniquely with a student in the NYC assignment records. Less than 0.45% of students with known assignments in the records of the NYC assignment system records could not be merged with a student in the DOE records. These students were not included in the analysis.

C.2 Applicant Sample Construction

Our goal is to consider first-time applicants to the NYC public (unspecialized) high school system who live in New York City and attend a public middle school in 8th grade. Below, we describe the procedure used to construct the samples. The selection procedure is also summarized in Table B1.

Welfare Sample

The welfare samples are constructed from the NYC DOE's records for all students enrolling in 9th grade at a high school in the academic years 2003-04 and 2004-05.

Because our choice set in the demand analysis will be restricted to unspecialized, non-charter high schools in the public school system, the welfare sample does not include students who matriculated to such schools.

Of the 92,623 8th grade students matriculating into 9th grade at a NYC public school in 2002-03, 11,790 (12.73%) students went to a private middle school and were dropped. Another 8,051 (8.69%) students were not included in the analysis because their assignments were unknown, or

because they matriculated to either a Specialized high school or a charter school. Finally, we exclude students in schools that were closed (i.e. no assignments in the new system).

In 2003-04, about 1.3% of students had also participated in the old mechanism, presumably because these students repeated 8th grade. These students were considered part of the 2002-03 sample and only their 2002-03 high school assignment is considered in our analysis. We also drop private middle school students and those not assigned to public school. These fractions were similar to the 2002-03 numbers, at 12.21% and 8.13% respectively. We also drop students who were assigned to new schools.

These selections into the sample leave us with 70,358 students in 2002-03 and 66,921 students in 2003-04. Students who may have been assigned to a high school program through a process other than the Main round are included in these samples.

Demand Sample

This sample is sourced from the NYC Assignment system's records on the participants in the Main round of the mechanism. As discussed in the text, we use data only from the Main round of the mechanism because this round has the most desirable incentive properties.

We do not want to exclude students on the basis of final assignment to avoid selecting on the choice to leave the public school system. In order to most closely match welfare sample construction, we select the demand sample only on characteristics that can be considered as exogenous at the time of participation.

Since we focus on first-time applicants in 8th grade, we exclude 747 students who were part of the 2002-03 files and 5,311 students who were 9th graders. Presumably, these students were held back in 8th or 9th grade. This leaves us with a sample of 81,297 8th grade students.

Of the 8th-grade participants, 9,301 (11.44%) students were from private middle schools and were dropped. We also excluded students designated as belonging to the top 2% of their middle school classes because these students are prioritized at education option schools, creating incentives to misreport their preferences. These are 2.5% of the non-private 8th grade population.

A total of 216 students did not rank any public schools in our sample. After excluding these students, a total of 69,907 students remain in the sample we use for the demand analysis.

C.3 Programs/Schools

NYC Department of Education School Report Cards

The school characteristics were taken from the report card files provided by the NYC Department of Education. These data provide information on a school's enrollment statistics, racial composition of student body, attendance rates, suspensions, teacher numbers and experience, and graduating class Regents Math and English performance. A unique identifier for each school allows these data to be merged with data from our other sources.

There were significant differences in the file formats and field names across the years. To keep the school characteristics constant across years, we use the data from the 2003-04 report cards as the primary source. This corresponds to information from school years prior to the new assignment mechanism. Except for data on the math and reading achievement, variable

descriptions were comparable across years. For these comparable variables, we used the 2002-03 data only when the 2003-04 data were not available. The coverage of the characteristics for the school sample is enumerated in Table B2.

Assignment System and DOE files

Assignment data contain a list of all school programs in the public school system along with an identifier for the associated high school. The DOE provided a separate file with data on the school addresses and identifiers that allows us to merge that information with the assignment system database. A second identifier can be used to merge these data with other fields in the department of education records described above.

Across the two years, the high school identifiers in the files were inconsistent for a small number of schools in our sample. These were matched by school name and address. One school moved from Brooklyn to Manhattan and was investigated to ensure the records were appropriately matched.

Program Characteristics

Program characteristics are taken from the DOE's High School Directory, which is made available to students before the application process. Reliable data on program types were not available in 2002-03. The 2002-03 program types were imputed from the 2003-04 program types if the program was present in both years. Otherwise, the program was categorized as a general program.

The numerous program types were aggregated into fewer broad categories. The items in the list below are the aggregated categories that include all the subcategories described by the data.

1. *Arts:* Dance, Instrument Performance, Musical Theater, Performing and Visual, Performing Arts, Theater, Theater Tech, Visual Arts, Vocal Performance.
2. *Humanities/Interdisciplinary*: Education, Humanities/Interdisciplinary.
3. *Business/Accounting*: Accounting, Business, Business Law, Computer Business, Finance, International Business, Marketing, Travel Business.
4. *Math/Science*: Engineering, Engineering – Aerospace, Engineering – Electrical, Environmental, Math and Science, Science and Math.
5. *Career*: Architecture, Computer Tech, Computerized Mech, Cosmetology, Journalism, Veterinary, Vision Care Technology.
6. *Vocational*: Auto, Aviation, Clerical, Construction, Electrical Construction, Health, Heating, Hospitality, Plumbing, Transportation.
7. *Government/law*: Law, Law Enforcement, Law and Social Justice, Public Service.
8. *Other:* Communication, Expeditionary, Preservation, Sports.
9. *Zoned*

10. *General*: General, Unknown.

Finally, some programs adopt a language of instruction other than English. We categorized the languages as Spanish, English, Asian Languages, and Other.

C.4 School Sample Construction

We consider NYC public middle school 8th grade students assigned to public high schools that are not charters, specialized or parochial. Our analysis uses two school samples, one for each year in our analysis.

To construct these samples, we started with the set of schools and programs in the assignment records. To analyze rank data, we added the set of school programs that were ranked by any student in our demand sample. This initial set consists of 743 (301) programs (schools) in 2002-03 and 677 (293) programs (schools) in 2003-04.

In 2003-04, this list contained 62 parochial school programs. We verified that each of the 130 students matriculating to these school programs were private middle-schoolers. These schools were dropped from the analysis because private middle-schoolers are not in the population of interest. Subsequently, we dropped all charter and specialized high school programs and other school programs that do not have assignments and were not ranked by any student in our sample.

A total of 9 continuing student programs accepted students only from their associated middle school. Since these programs cannot be chosen by students who were not in those schools in 8th grade, we combine these programs with a generic program (e.g., unscreened, English, general/humanities/math). Rank-order lists for students who ranked both the continuing students' only program and the associated program were modified as described below.

Finally, we dropped new and closed schools from the analysis. Closed schools were ones that admitted students in 2002-03 but not in 2003-04. The set of new schools was collected from a separate DOE directory of new schools. These schools were not well advertised and very few students ranked them, making calculations with those schools unreliable.

The number of schools and programs at each stage of our selection procedure is also summarized in Table B2.

C.5 Program Capacities

Program capacities are not provided separately in the data files. We have estimated program capacities from the actual match files and students' final assignments. The capacity of each program is initially set to zero. If a student in our demand sample is assigned a program at the end of the assignment process, the capacity of the program is increased by one. Otherwise, if the student is assigned a program in the Main round, the capacity of the program is increased by one. Finally, if a student is not assigned in the Main round and is assigned a program in the Supplementary round, the capacity of the program is increased by one.

Education Option programs are divided into six buckets: High Select, High Random, Middle Select, Middle Random, Low Select and Low Random. Bucket capacities are calculated as above by taking into account the category of the assigned student. For example, if a student of High category is assigned an Education Option program, then the capacity of a High Select bucket is

increased by one. If the current capacity of the High Select bucket is less than or equal to that of High Random, then the capacity of the High Select bucket is increased, otherwise the capacity of the High Random bucket is increased.

C.6 Program Priorities

Program type determines how students are priority-ordered. The data contains a list of all programs with program-specific information, including type, building number, street address, etc. When students have the same priority, the tie is broken randomly. The random numbers are generated by computer during our simulations.

The assignment data contains several fields that determine a student's priority order in programs. Priority group is a number assigned by the NYC DOE depending on students' home addresses, program location, etc. High school rank is a number assigned by each program. This number may reflect a student's ranking among all applicants to an Education Option program, whether a student attended the information session of a limited unscreened program, etc. These fields are provided for every program every student ranked. Students applying to Educational Option programs are placed into one of three categories based on their score on the 7th grade reading test: top 16 percent (high), middle 68 percent (middle), and bottom 16 percent (low). Student categories are included in the assignment data.

Unscreened programs order students based on their random numbers only. Limited unscreened and formerly zoned programs order students first by priority group and then by random number within the priority group. Screened programs order students by priority group and then by high school rank. Each Education Option program orders all applicants for each of six buckets, High Select, High Random, Middle Select, Middle Random, Low Select and Low Random. A High Bucket orders high category students first, then middle category students, then low category students. A Middle Bucket orders middle category students first, then high category students, then low category students. A Low Bucket orders low category students first, then high category students, then middle category students. A Select bucket orders students within each category by priority order and then by high school rank. A Random bucket orders students within each category by priority order.

C.7 Regents Test and Graduation Outcomes

Regents

The NYC Regents test file contains the date and raw score for each tested student from 2004 to 2010. Regents exams are mandatory state examinations on which performance determines whether a student is eligible for a Regents high school diploma in New York. There are Regents examinations in English, Global History, US History, and multiple exams in Math and Science. A Regents exam typically has a multiple choice section and a long answer or essay component, and each exam usually lasts for three hours. The English exam, however, consists of two three-hour pieces over two days. The exam has a locally-graded component and Dee, Jacob, McCrary, and Rockoff (2016) illustrate how test scores bunch near performance thresholds.

The New York State Board of Regents governs and designs the Regents exams. Starting in

2005, they started to modify the math exams. At the beginning of our sample, the two math exams were Elementary Algebra and Planar Geometry (Math A) and Intermediate Algebra and Trigonometry (Math B). Two new math exams, Integrated Algebra I (Math E) and Geometry (Math G), have since been phased in. Since students typically either take Math A or Math E, we focus on the score on the test taken first, taking the Math A score when both are contemporaneous. There are Regents science exams in Earth Science, Living Environment, Chemistry, and Physics. The science outcome we focus on is Living Environment because it is the most commonly taken Regents science exam. English and US History Regents exams are typically taken in 11th grade.

The Regents file does not have the test date, and instead only has a variable indicating the term ("termcd"). Based on discussions with the DOE, we convert term to fall if the termcd is "1", "5", "a", or "A" and to spring if the termcd is "2", "3", "4", "6", or "7." The DOE indicated exceptions at the following school DBNs where the termcd of "2" refers to the fall semester: 79M573, 79M612, 32K564, 02M560, 10X319, 02M575, 22K585, 12X480, 03M505, 02M570, 21K525, 21K540, 19K409, 17K489, 15K698, 14K454, 14K640, 07X379, 11X265, 15K529, 08X377, 05M285, 21K728, 02M303, 25Q792, 18K578, 24Q520, and 19K431. If the student takes a subject before 9th grade, that subject is dropped for that student. If a student takes the test more than once after 9th grade, we used the test score from the earliest date. There are a small number of cases where there is more than one score on the same date, and this date is the first date after entering 9th grade. In some of these cases, there are two different test codes, where one code ends with a "2". We used the score corresponding to the test that does not end with a "2". Otherwise, we treated the score as missing.

We focus on the results in the Mathematics and English tests. Given the existence of multiple Math tests we take the earliest test between Math A and Math B, which are the two most common tests. If a student takes both tests in the same school year and term we use the Math A result.

For each subject, we standardized scores to have mean zero and standard deviation within subject for each cohort of test-takers by year and test time.

Graduation

The Graduation file contains the discharge status of all public school students from 2005 to the Spring 2012. For application cohorts 2002-03 and 2003-04, students should start school in Fall 2003 and Fall 2004 and graduate on-time in Spring 2007 and Spring 2008, respectively. To code graduation type, we use the following discharge codes: a) 26, 30, and 61 (discharge codes for a local diploma)?b) 27, 46, and 60 (discharge codes for a Regents diploma) c) 28, 47, and 62 (discharge codes for an Advanced Regents diploma).

It is possible for students to graduated with several different discharge files. We classify students as having received or not any diploma (local or Regents) and we distinguish among students who received any Regents diploma or not.

C.8 Miscellaneous Issues

Modifications to the rank-order list

1. In the Main round, some students ranked programs that were either charter schools or Specialized High Schools. These programs are not in the sample of schools we consider and were likely ranked by in error the students. In such cases, programs were removed from the rank-order lists and rank-order lists were made contiguous where all programs ranked below a program not in the sample were moved up in the rank-order lists. These programs were observed a total of 795 times in the data. Thirty students ranked only charter or specialized programs.

2. The rank-order lists of students who ranked continuing student programs were modified as follows: First, the lists of all students who ranked only the continuing student programs were modified so that the student ranked the associated generic program instead. When students ranked both the generic program and the associated continuing student program, the list was modified so that only the associated program was ranked, and it was ranked at the highest of the two ranked positions. All programs ranked at positions below the lower-ranked of the two programs were moved up by one. A total of 46 students ranked both the continuing program and the generic program, to which we mapped the continuing program to. In 17 cases, these ranks were not consecutive.

Table C1. Student Sample Selection

	Mechanism Comparison		
	Uncoordinated	Coordinated	
	2002-2003	2003-2004	2004-2005
	(1)	(2)	(3)
Number of students in the NYC DOE student file	100,669	97,569	96,327
Number of students in the rank data			
Excluding students in both 2002-03 and 2003-04 files from 2003-04		96,275	
Excluding 9th grade students	92,623	89,062	90,250
Excluding private middle school students	80,833	78,183	80,093
Excluding students with addresses outside the five boroughs	80,725	78,089	79,977
Total number of students with known assignments to sample schools	75,515	73,989	75,049
Excluding students attending specialized high schools	72,725	70,992	71,861
Excluding students attending charter schools	72,681	70,886	71,749
Excluding students in closed and new schools	70,358	66,921	69,013
Excluding top 2% students			
Excluding students that did not rank any sample schools			

Notes: Uncoordinated mechanism refers to the 2002-03 mechanism and coordinated mechanism refers to the 2003-04 mechanism based on deferred acceptance. A stud is missing, cannot be geocoded, or places the student outside of New York City. A distance observation is invalid if it is missing or is greater than 65 miles.

Table C2. Construction of School Sample

	Uncoordinated		Coordinated			
	2002-2003		2003-2004		2004-2005	
	Programs	Schools	Programs	Schools	Programs	Schools
	(1)	(2)	(3)	(4)	(5)	(6)
Programs where NYC public school students assigned	743	301	669	293	658	322
Adding additional programs ranked by students			677	294	764	338
Excluding parochial schools	681	239	677	294	752	331
Excluding specialized schools	669	232	665	287	750	329
Excluding charter schools	667	230	663	285	702	315
Excluding programs with no assignments or ranking	637	225	648	284	691	313
Combining continuing education programs	637	225	639	284	691	313
Excluding closed schools	612	215	639	284	691	313
Excluding schools opened after HS directory printed*	612	215	558	235	661	283
Programs/schools ranked by students in sample			497	234	660	283

Notes: 13 continuining student programs were merged with a generic program at host school. Parochial schools in 2002-03 only have private middle school students assigned to them and are not ranked by students in the demand sample. *A total 20 schools and 23 programs opened before HS directory printed are included in 2003-04.

Table C3. Coverage of School Characteristics

	Uncoordinated 2002-03 (1)	Coordinated 2003-04 (2)	Both Years (3)
Total number of schools in the sample	215	234	215
9th grade enrollment	196	199	189
Race	196	199	189
Attendance Rate	196	199	189
Percent Free Lunch	196	198	189
Percent of teachers less than 2 years experience	219	223	212
High Math Achievement	198	200	191
High English Achievement	180	177	173
Percent Attending College	171	167	165

Notes: Table reports the number of schools with the characteristics from New York State Report cards.