(NON)RANDOMIZATION: A THEORY OF QUASI-EXPERIMENTAL EVALUATION OF SCHOOL QUALITY

By

Yusuke Narita

November 2016
Revised August 2017

COWLES FOUNDATION DISCUSSION PAPER NO. 2056R

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

(Non)Randomization:
A Theory of Quasi-Experimental Evaluation of School Quality

Yusuke Narita*†

August 19, 2017

Abstract

Many centralized school admissions systems use lotteries to ration limited seats at oversubscribed schools. The resulting random assignment is used by empirical researchers to identify the effect of entering a school on outcomes like test scores. I first find that the two most popular empirical research designs may not successfully extract a random assignment of applicants to schools. When do the research designs overcome this problem? I show the following main results for a class of data-generating mechanisms containing those used in practice: One research design extracts a random assignment under a mechanism if and practically only if the mechanism is strategy-proof for schools. In contrast, the other research design does not necessarily extract a random assignment under any mechanism.

Keywords: Matching Market Design, Natural Experiment, Program Evaluation, Random Assignment, Quasi-Experimental Research Design, School Effectiveness

*Yale University, Department of Economics and Cowles Foundation. Email: `yusuke.narita@yale.edu`

†I am grateful to discussions with Atila Abdulkadiroğlu, Daron Acemoglu, Nikhil Agarwal, Josh Angrist, Nick Arnosti, Eduardo Azevedo, Ian Ball, Dirk Bergemann, Yeon-Koo Che, David Deming, Peter Hull, Michihiro Kandori, Fuhito Kojima, Parag Pathak, Debraj Ray, Jaehee Song, Jean Tirole, Chris Walters, Kohei Yata, and seminar participants at Yale, Columbia, MIT, Seoul, Tokyo, Kyoto, Osaka, and Otaru.

1 Introduction

The spread of choice in public education is giving more families the option to attend a school other than their neighborhood default. As choice has proliferated, school assignment has grown increasingly centralized and algorithmic in order to respect heterogeneous preferences and various priorities based on family background. Centralized assignment mechanisms solve the problem of matching the demand for school seats with their limited supply by using centralized algorithms. Such mechanisms are employed in numerous school and college admission institutions in America, Africa, Asia, and Europe. Well-designed centralized assignment provides a transparent way to achieve a fair and efficient school seat allocation, while narrowing the scope for strategic behavior (Abdulkadiroğlu and Sönmez, 2003).

Moreover, centralized assignment generates valuable data for empirical research on education. In particular, when a school is oversubscribed, mechanisms often use random lotteries to ration limited seats among applicants. This generates quasi-experimental variation in school assignment that opens the door to a variety of impact evaluations. Researchers used such variation to study schools in the Bay Area (Bergman, 2016), Boston (Angrist et al., 2016), Charlotte-Mecklenburg (Hastings et al., 2009; Deming, 2011; Deming et al., 2014), Denver (Abdulkadiroğlu et al., 2017), and New York (Bloom and Unterman, 2014; Abdulkadiroğlu et al., 2014b).[1]

Centralized assignment mechanisms combine lotteries, preferences, and priorities into complex stratified randomized experiments. Empirical research designs based on such mechanisms therefore need to condition on appropriate objects to isolate random components of their data-generating mechanisms. Yet, the above empirical work provides only a limited foundation for how the research designs extract a conditionally random assignment.[2]

This paper studies when widely-used empirical research designs successfully extract conditionally random assignment of students to schools. I focus on the two most popular research designs. These designs are applicable to any centralized mechanism that assigns students to

[1]See also Hastings et al. (2012). Other studies use related regression-discontinuity-style tie-breaking rules to evaluate college majors in Norway (Kirkeboen et al., 2016) and in Chile (Hastings et al., 2013), daycare in Italy (Fort et al., 2016), privately managed public schools in Trinidad and Tobago (Beuermann et al., 2016), as well as popular selective schools in Ghana (Ajayi, 2013), Kenya (Lucas and Mbiti, 2014), Romania (Pop-Eleches and Urquiola, 2013), Trinidad and Tobago (Jackson, 2010, 2012), and the U.S. (Abdulkadiroğlu et al., 2014a; Dobbie and Fryer, 2014). Narita (2015) uses lottery-based randomization to identify a structural model of evolving demand for schools.

[2]An exception is Abdulkadiroğlu et al. (2017). See the literature review at the end of this introduction for the relationship between my paper and theirs. The other papers simply check empirical "covariate balance." That is, they compare the treatment and control groups by baseline characteristics or covariates that are fixed at the time of treatment assignment and not used for it. If the two groups' covariates are similar (covariates are balanced), it is interpreted as not rejecting randomization. Covariate balance is necessary but not sufficient for randomization.

schools by combining: (1) applicants' rank-ordered preferences over schools, (2) applicants' priority statuses (e.g., walk zone) at schools, and (3) lottery numbers for breaking ties in priority status. Each of the empirical examples above uses one of these research designs.

The first research design is what I call the *first-choice research design*. This design focuses on applicants who rank a given treatment school first and are in the "marginal priority" group at the school. The marginal priority group means a priority group where some students are assigned to the treatment school while others are not. Within this *first-choice subsample*, some applicants are assigned to the treatment school while others are not, though all students rank the treatment school first and share the same priority. Thus it appears that solely lottery numbers determine treatment assignment. Based on this idea, the first-choice research design assumes that applicants are randomly assigned to or rejected by the treatment school conditional on being in the first-choice subsample. In the first-choice subsample, the analyst then compares the outcomes (e.g., test scores) of students who are assigned to the treatment school against those who are not assigned. The outcome difference between the two groups is interpreted as a causal effect of the treatment school.

Despite its intuitive construction, it turns out that the first-choice research design may not extract a random assignment in general. That is, applicants in the first-choice subsample may not share the same assignment probability at the treatment school.[3] This motivates me to investigate the conditions under which the first-choice design extracts a random assignment. I provide such conditions for a class of data-generating mechanisms nesting those used in the above empirical examples: The first-choice research design extracts a conditionally random assignment for a mechanism if and practically only if the mechanism is strategy-proof for schools.[4]

This result has important implications for applied research. It justifies the first-choice research design for mechanisms that are known to be strategy-proof for schools, such as the Boston (immediate acceptance) mechanism (Ergin and Sönmez, 2006). My result also suggests that attention should be paid to the research design for other widely-used mecha-

[3]How can the first-choice design fail to extract a random assignment? To gain intuition, imagine the treatment school A has only one seat, and the first-choice subsample contains two students, 1 and 2. Student 1 ranks only A while 2 ranks other schools below A. When 2 has a better lottery number than 1, 1 is rejected by A and stops applying since 1 ranks only A. When 1 has a better lottery number, 2 is rejected by A and then applies for other schools, potentially crowding out other students there. These crowded-out students may apply for A, which may crowd student 1 out of A. Such chain reactions of rejections and new applications dilute 1's, but not 2's, treatment assignment probability at A. As a result, 1 and 2 may have different treatment assignment probabilities even though these students constitute the first-choice subsample. This prevents the first-choice design from extracting a random assignment. Section 3.1 provides a more precise example.

[4]The if part is exactly true. The practically-only-if part means that the first-choice design sometimes fails to extract a random assignment for any non-strategy-proof mechanisms used in the above empirical studies (but not for all possible non-strategy-proof mechanisms).

nisms that are not strategy-proof for schools, such as the deferred acceptance mechanism, a mechanism used in Charlotte, and the top trading cycles mechanism.

By contrast to the above partial justification for the first-choice design, no similar sufficient condition is obtainable for another popular research design. I call this alternative the *qualification instrumental variable (IV) research design*. Unlike the first-choice design (trying to make assignments random by focusing on a subset of students), the qualification IV design considers all students. It then codes a supposedly random instrumental variable for non-random assignments. The IV is based on "qualification," i.e., whether a student's lottery number is better than the worst number offered a seat at the treatment school (conditional on priority).

I find that even in the simple case with no priorities and unit school capacities, the qualification IV research design does not necessarily extract a random assignment for any mechanism (within my mechanism class); that is, applicants may not share the same conditional probability of qualification at the treatment school. This shows a contrast between the qualification IV design and the first-choice design, as summarized in Table 1.[5]

Table 1: Summary of the main results

Do empirical research designs always extract a random assignment?	1st choice research design	Qualification IV research design
Under mechanisms strategy-proof for schools (e.g., Boston mechanism)	✓	×
Under other mechanisms (e.g., deferred acceptance, Charlotte, and top trading cycles mechanisms)	×	×

Before I move on to the analysis, two remarks are in order about the initial result using strategy-proofness for schools. First, I do *not* assume that schools have preferences or are strategic in reality. This is because my analysis treats strategy-proofness not as a desideratum or incentive compatibility constraint but rather as an algorithmic property, which turns out to mathematically imply the success of an empirical research design. Therefore, the empirical implications of my result are free from any assumption about school behavior or preference.

[5]This comparison is based on under which mechanisms each design always extracts a random assignment. The same point can be made even if I fix a particular mechanism: The above result for the qualification IV implies that for any mechanism, the qualification IV may not extract a random assignment. By contrast, the first-choice design always extracts a random assignment under several mechanisms, as shown below. Therefore, even conditional on a particular mechanism, the first-choice design is weakly more likely to extract a random assignment than the qualification IV design.

In addition, I need no assumption on student behavior (e.g., truthful preference reporting). I study whether assignment algorithms do or do not extract a random assignment conditional on any reported preferences and without reference to true preferences. As a result, my results do not depend on whether the reported preferences are truthful or not.

Second, the initial result — strategy-proofness for schools is sufficient for the first-choice design to extract a random assignment — has an additional empirical implication. Particularly, it provides an asymptotic support for the first-choice design even for mechanisms that are not strategy-proof in general. This is because such non-strategy-proof mechanisms like deferred acceptance are known to be approximately strategy-proof for schools in certain large markets with many students and schools (Roth and Peranson (1999) and subsequent studies).[6] This may explain why the first-choice design appears to extract a random assignment in empirical applications even for non-strategy-proof mechanisms. Viewed differently, the existing empirical justification for the first-choice design (in the form of covariate balance regressions) may suggest the empirical relevance of theoretical results on strategy-proofness in large markets.

The rest of this paper is organized as follows. After a literature review, the next section introduces my model. Section 3 defines the first-choice research design and gives conditions under which the research design extracts a random assignment. Section 4 analyzes the alternative qualification IV design and compares it with the first-choice design. Section 5 confirms that my results are robust to a variety of modifications to the definitions of research designs and randomization. Finally, Section 6 summarizes the empirical implications of my theoretical results and suggests an agenda for further research.

Related Literature

This paper theoretically studies the empirical practice in econometric evaluations of school effectiveness, such as Hastings et al. (2009); Deming (2011); Deming et al. (2014); Bloom and Unterman (2014); Abdulkadiroğlu et al. (2014b); Bergman (2016); Angrist et al. (2016). My analysis reveals the connection between their empirical strategies and theoretical market design studies, especially those on strategy-proofness (Ergin and Sönmez, 2006; Roth and Peranson, 1999; Immorlica and Mahdian, 2005; Kojima and Pathak, 2009; Azevedo and Budish, 2013; Lee, 2016; Ashlagi et al., 2016). On top of them, Abdulkadiroğlu et al. (2017) is closely related. They develop a large-sample framework based on an asymptotic approximation assuming a growing number of students and school seats. They use their model to propose an improvement over the first-choice and qualification IV designs and apply the

[6]See, among others, Immorlica and Mahdian (2005); Kojima and Pathak (2009); Azevedo and Budish (2013); Lee (2016); Ashlagi et al. (2016).

improved design to evaluate charter schools in Denver. They also confirm that the first-choice and qualification IV designs extract a random assignment for many mechanisms in the limit of their large market sequence. In contrast, the current paper allows for general finite markets and provides conditions for the first-choice or qualification IV design to extract a random assignment in a finite sample. These conditions allow me to compare the two research designs, as in Table 1. I also provide additional large market justifications for the first-choice design in large market models different from Abdulkadiroğlu et al. (2017)'s.

2 Framework

I use a model of school-student assignment with coarse school priorities and lotteries. There are a finite set I of students and a finite set S of schools. Each student $i \in I$ has a strict preference $\succ_i$ over $S \cup \{\emptyset\}$, where $\emptyset$ denotes the outside option of the student. This $\succ_i$ is i's reported preference recorded in the data; I do not make any assumption about whether $\succ_i$ is truthful or not. School s is said to be acceptable for student i if $s \succ_i \emptyset$. A preference profile for all students is denoted by $\succ_I \equiv (\succ_i)_{i\in I}$. Each school s has a capacity $c_s \in \mathbb{N}$ where $\mathbb{N}$ is the set of positive integers. Schools also grant students coarse priorities. $\rho_{is} \in \{1, ..., K\}$ denotes student i's priority at school s where $\rho_{is} < \rho_{js}$ means s prioritizes i over j. Motivated by public school applications, I assume every student is acceptable to every school. The number of possible priority statuses K may change as the number of students $|I|$ changes. Priorities may be coarse in the sense that it is possible that $\rho_{is} = \rho_{js}$ for some $i \neq j$. Let $\rho_s \equiv (\rho_{is})_{i\in I}$ and $\rho \equiv (\rho_s)_{s\in S}$. Denote the **type** of student i by $\theta_i \equiv (\succ_i, (\rho_{is})_{s\in S})$. I call $X \equiv (I, S, \succ_I, (c_s)_{s\in S}, (\rho_s)_{s\in S})$ an **assignment problem**.

2.1 Generalized Deferred Acceptance Mechanisms

A (stochastic) mechanism maps each assignment problem into a distribution over matchings between students and schools. Mechanisms usually use lotteries to break ties in priority and then use the resulting strict priorities to create a matching. A random variable R_{is} denotes student i's lottery number at school s. Assume that at each school, R_{is} is iid across students according to $U[0, 1]$. For the correlation of lottery numbers across different schools, I consider two focal regimes. Under a "single tie breaker" (STB), each student has a single lottery number used by all schools, i.e., $R_{is} = R_{is'}$ always holds for all i, s, and s'. Under a "multiple tie breaker" (MTB), each student has an independent lottery number at each school, i.e., R_{is} and $R_{is'}$ are independent for all i and $s \neq s'$.[7] Let $r_{is} \in [0, 1]$ denote

[7] In reality, most school districts use STB, though some cities like Washington, D.C., New Orleans, and Amsterdam use MTB. It is possible but requires messier notation to extend my analysis to any structure in

i's realized lottery number at school s and let $R \equiv (R_{is})_{i\in I, s\in S}$, and $r \equiv (r_{is})_{i\in I, s\in S}$. When referring to any realized lottery number vector r, I assume no tie and $r_{is} \neq r_{js}$ for all students i, j, and school s.

To define mechanisms of interest, I first introduce the following (student-proposing) **deferred acceptance (DA) algorithm** (Gale and Shapley, 1962). The DA algorithm produces an assignment by using any given strict student preferences and strict school priority orders as follows.

- Step 1: Each student i applies to her most preferred acceptable school (if any). Each school tentatively keeps the highest-ranking students up to its capacity, and rejects every other student.

In general, for any subsequent step $t \geq 2$,

- Step t: Each student i who was not tentatively matched to any school in Step $t-1$ applies to her most preferred acceptable school that has not rejected her (if any). Each school tentatively keeps the highest-ranking students up to its capacity from the set of students tentatively matched to this school in previous step $t-1$ and the students newly applying. The school rejects every other student.

The algorithm terminates right after the first step at which no student applies to any school. Each student tentatively kept by a school at that step is allocated a seat in that school, resulting in an assignment. I use this algorithm to define a class of mechanisms of interest.[8]

Definition 1. A **generalized deferred acceptance (gDA) mechanism** φ is a mechanism that can be expressed as the following procedure. Take any assignment problem as given.

(1) Draw lottery numbers r according to its lottery regime (STB or MTB).

(2) For each student i and school s, compute the modified priority

$$\rho^{\varphi}_{is} \equiv f^{\varphi}(\rho_{is}) + g^{\varphi}(rank_{is}),$$

where $f^{\varphi} : \mathbb{N} \to \mathbb{N}$ is a strictly increasing function, $g^{\varphi} : \mathbb{N} \to \mathbb{N}$ is a weakly increasing function, and $rank_{is}$ is the preference rank of school s in student i's preference $\succ_i$. For example, $rank_{is} = 2$ if s is i's second choice school. Define school s's ex post strict modified priority order $\succ^{\varphi}_{r_s}$ over students by $i \succ^{\varphi}_{r_s} i'$ if $\rho^{\varphi}_{is} + r_{is} < \rho^{\varphi}_{i's} + r_{i's}$.[9]

between STB and MTB where some schools use a common lottery while others use independent ones.

[8]Others also use similar classes of mechanisms. See, for example, Ergin and Sönmez (2006); Pathak and Sönmez (2008); Agarwal and Somaini (2015); Abdulkadiroğlu et al. (2017).

[9]Note that lottery number r_{is} is used only for tie-breaking in modified priority ρ^{φ}_{is} since ρ^{φ}_{is} is an integer and r_{is} is in $(0, 1)$.

(3) Given $\succ_I$ and $(\succ^{\varphi}_{r_s})_{s \in S}$, run the DA algorithm to produce an assignment, where each school s's priority order is given by $\succ^{\varphi}_{r_s}$.

gDA mechanisms are parametrized by the lottery regime (STB or MTB) and the priority modification function $(f^{\varphi}, g^{\varphi})$. This gDA class includes most of the mechanisms used in empirical research as I now explain.

Deferred Acceptance Mechanism

Given an assignment problem and realized lottery numbers, the **deferred acceptance (DA) mechanism** (Gale and Shapley, 1962; Abdulkadiroğlu and Sönmez, 2003) makes a matching through the DA algorithm in which schools' strict priorities are induced by $\rho_{is} + r_{is}$. The DA mechanism makes no modification to priorities and corresponds to the gDA mechanism with $f^{\varphi}(m) = m$ and $g^{\varphi}(n) = 0$.

Boston (Immediate Acceptance) Mechanism

The **Boston (immediate acceptance) mechanism** (Abdulkadiroğlu and Sönmez, 2003; Ergin and Sönmez, 2006) is defined through the following immediate acceptance algorithm.

- Step 1: Each student i applies to her most preferred acceptable school (if any). Each school accepts its highest-priority (with respect to $\rho_{is} + r_{is}$) students up to its capacity and rejects every other student.

In general, for any step $t \geq 2$,

- Step t: Each student who has not been accepted by any school applies to her most preferred acceptable school that has not rejected her (if any). Each school accepts its highest-priority (with respect to $\rho_{is} + r_{is}$) students up to its remaining capacity and rejects every other student.

The algorithm terminates immediately after the first step in which no student applies to any school. Each student accepted by a school at some step of the algorithm is allocated a seat in that school. The immediate acceptance algorithm differs from the DA algorithm in that when a school accepts a student at a step, in the immediate acceptance algorithm, the student is guaranteed that school, while in the deferred acceptance algorithm, that student may be later displaced by another student with a better priority status.

The Boston mechanism can be interpreted as modifying priorities so that each school prioritizes students ranking it higher over students ranking it lower. It is known that the Boston mechanism is a gDA mechanism with $f^{\varphi}(m) = m$ and $g^{\varphi}(n) = (K+1)n$ (Ergin and

Sönmez, 2006). Under this $(f^{\varphi}(m), g^{\varphi}(n))$, any school's modified priority order induced by ρ^{φ}_{is} is lexicographic in preference ranks and priority statuses. That is, $i \succ^{\varphi}_{r_s} i'$ for all i and i' with $rank_{is} < rank_{i's}$ regardless of the original priorities ρ_{is} and $\rho_{i's}$ and lottery numbers r_{is} and $r_{i's}$; $i \succ^{\varphi}_{r_s} i'$ for all i and i' with $rank_{is} = rank_{i's}$ and $\rho_{is} < \rho_{i's}$ regardless of lottery numbers r_{is} and $r_{i's}$.

Charlotte Mechanism

The mechanism used in Charlotte is the same as the Boston mechanism except that each school respects the walk zone priority ahead of preference ranks so that every student is guaranteed a seat at her walk zone school (Hastings et al., 2009; Deming, 2011; Deming et al., 2014). Assume without loss of generality that $\rho_{is} = 1$ means i has walk zone priority at s. The **Charlotte mechanism** is a gDA mechanism with $f^{\varphi}(m) = m + 1\{m > 1\}[K + (K+1)|S|]$ and $g^{\varphi}(n) = (K+1)n$. Under this $(f^{\varphi}(m), g^{\varphi}(n))$, any school's modified priority order is lexicographic in the walk zone priority status, preference ranks, and other (non-walk-zone) priority statuses.[10]

3 First-Choice Empirical Research Design

As explained in the introduction, many empirical studies use data from gDA mechanisms to identify and estimate the causal effect of assignment to a treatment school on outcomes such as test scores, crime rates, college attendance, and earnings.[11] Their empirical research designs fall into two categories. I start with analyzing one of them and move on to the other in Section 4.

To describe the first empirical strategy, fix any gDA mechanism φ and assignment problem X that generates the data at hand. Following the standard notation in econometrics, let $D_{is}(r) = 1$ if student i is assigned the treatment school s under (realized or counterfactual) lottery number profile r; $D_{is}(r) = 0$ otherwise. I consider the set of students who rank s first and are in s's "marginal priority group," where some students are assigned s but others

[10]That is, $i \succ^{\varphi}_{r_s} i'$ for all i and i' with $\rho_{is} = 1$ and $\rho_{i's} > 1$; $i \succ^{\varphi}_{r_s} i'$ for all i and i' with $1\{\rho_{is} = 1\} = 1\{\rho_{i's} = 1\}$ and $rank_{is} < rank_{i's}$; $i \succ^{\varphi}_{r_s} i'$ for all i and i' with $1\{\rho_{is} = 1\} = 1\{\rho_{i's} = 1\}$, $rank_{is} = rank_{i's}$, and $\rho_{is} < \rho_{i's}$.

[11]Many of the studies mentioned in the introduction investigate the effect of a group of schools rather than an individual school. My analysis extends to such group-level treatments too. Also, when the effect of interest is that of attendance or enrollment rather than assignment, the analyst would see attendance or enrollment as the endogenous treatment and use assignment as an instrument for the treatment. The analyst would then use an instrumental variable method to estimate the effect of attendance or enrollment. My analysis is applicable to such instrumental variable settings. See footnote 15.

are not though all of them share the same priority at s. That is, define

$$
\begin{aligned}
&First_s(r) \\
&\equiv \{i \in I | rank_{is} = 1 \text{ and } \exists i' \text{ such that } rank_{i's} = 1, \rho_{is} = \rho_{i's}, \text{ and } D_{is}(r) \neq D_{i's}(r)\}.^{12}
\end{aligned}
$$

Let r_0 be the realized profile of lottery numbers in the data.

The first widespread empirical strategy, which I call the first-choice research design, compares the outcomes of students with $D_{is}(r_0) = 1$ against those with $D_{is}(r_0) = 0$ within $First_s(r_0)$.[13] The outcome difference between the two groups is then interpreted as the causal effect of being assigned to school s for students in $First_s(r_0)$. The idea is that since all students in $First_s(r_0)$ rank s first and share the same priority at s, whether they get an offer from s should be determined solely by their lottery numbers and hence independent of students' covariates or choices potentially correlated with outcomes. Therefore, offers from s within $First_s(r_0)$ are thought of as being randomly assigned in a randomized controlled trial.

Albeit intuitive, for the first-choice research design to identify a causal effect by this logic, assignments to s within $First_s(r_0)$ have to be indeed random and not confounded by non-random preferences or priorities. This requirement is formalized as the following concept.

Definition 2. The first-choice research design extracts a random assignment for a gDA mechanism φ if for any assignment problem X, any school s, any potential lottery realization r, and any students $j, k \in First_s(r)$,

$$P(D_{js}(R) = 1) = P(D_{ks}(R) = 1).$$

An equivalent requirement is

$$P(D_{is}(R) = 1 | i \in First_s(r), \theta_i = \theta) = P(D_{is}(R) = 1 | i \in First_s(r)),$$

[12] Since $rank_{is} = rank_{i's} = 1$ holds and $f^{\varphi}(\cdot)$ is strictly increasing by definition, $\rho_{is} = \rho_{i's}$ is equivalent to

$$(f^{\varphi}(\rho_{is}) + g^{\varphi}(1) \equiv) \rho^{\varphi}_{is} = \rho^{\varphi}_{i's} (\equiv f^{\varphi}(\rho_{i's}) + g^{\varphi}(1)).$$

I could therefore replace $\rho_{is} = \rho_{i's}$ with $\rho^{\varphi}_{is} = \rho^{\varphi}_{i's}$ in the definition of $First_s(r)$ without changing anything in the following analysis. Note also that it is possible $First_s(r) = \emptyset$ for some or even all r.

[13] Applications of the first-choice research design include Hastings et al. (2009); Deming (2011); Deming et al. (2014); Abdulkadiroğlu et al. (2014b); Bloom and Unterman (2014); Angrist et al. (2016). The first three studies use data from the Charlotte mechanism while the remaining studies are based on the DA mechanism.

for any student type θ for which the left-hand-side conditional probability is well-defined. $P(D_{is}(R) = 1 | i \in First_s(r), \theta_i = \theta)$ means the probability of assignment to s for an arbitrary student of type θ in $First_s(r)$.

This property requires that conditional on being in $First_s(r_0)$, offers from s are random and independent of students' preferences and priorities summarized by θ_i. In the econometric terminology, this requires that the propensity score (Rosenbaum and Rubin, 1983) is constant across all students in $First_s(r_0)$.[14] Only under this conditionally random assignment are the treatment and control groups in $First_s(r_0)$ comparable with each other. Econometric program evaluation methods require this conditional independence for the first-choice research design to identify a causal treatment effect (Heckman and Vytlacil (2007) chapters 8 and 9, Manski (2008) chapters 3 and 7, Angrist and Pischke (2009) chapter 3.2).[15]

3.1 Motivating Example

While the first-choice research design is intuitive, this design may fail to extract a random assignment. Consider the following example.

Example 1. There are applicants 1, 2, 3, and schools A and B with the following preferences and priorities:

$$\succ_1 : A, B, \emptyset$$
$$\succ_2 : A, \emptyset$$
$$\succ_3 : B, A, \emptyset$$
$$\rho_A : 3, \{1, 2\}$$
$$\rho_B : 1, \{2, 3\},$$

where $\succ_1$: $A, B, \emptyset$ means 1 prefers A over B and both schools are acceptable for 1. ρ_A : $3, \{1, 2\}$ means that A prioritizes 3 over 1 and 2 and is indifferent between 1 and 2. The capacity of each school is 1. The treatment school is A.

[14] It is possible to define random assignment conditional on being in random $First_s(R)$, where R are random lottery numbers and $First_s(R)$ is a random set. Alternatively, it is also reasonable to define random assignment as that all students who rank school s first and who have the same priorities at school s share the same assignment probability at s. My result is robust to using such alternative definitions; see Section 5.1 for more discussions.

[15] When assignment within $First_s(r_0)$ is used as an instrument for an endogenous treatment such as enrollment, Definition 2 is interpreted as a conditional independence requirement for the instrument. For the instrument to identify a causal effect, it usually needs to additionally satisfy properties such as "exclusion" or "monotonicity." See Heckman and Vytlacil (2007); Manski (2008); Angrist and Pischke (2009). In this case, Definition 2 becomes a necessary condition for legitimate causal inference. See also Sections 4 and 5.2 for related discussions.

In this example, the first-choice research design does not extract a random assignment for A for the DA mechanism (with no priority modification). Under the DA mechanism, 1 is assigned to A when 1 has a better lottery number than 2 at A. Otherwise, 3 is assigned to A. Each of the two cases occurs with equal probability 0.5. Thus,

$$P(D_{1A}(R) = 1) = 0.5 \neq 0 = P(D_{2A}(R) = 1),$$

despite

$$First_A(r) = \begin{cases} \{1, 2\} & \text{if } r_{1A} < r_{2A} \\ \emptyset & \text{otherwise.} \end{cases}$$

Therefore, the first-choice research design does not extract a random assignment for the DA mechanism.

It is possible to create a similar counterexample even when there are no priorities (as long as ties are broken by MTB). Also, the first-choice research design may fail even if I modify it to the more conditional version that pools applicants who rank the treatment school first and share the same priority at *every* school. Finally, the problem with the first-choice design does not depend on short preferences; the problem turns out to persist even if I require every student to rank all schools. I explain these points in Section 5.2.

The above problem may bias treatment effect estimates. Imagine that school A has no real treatment effect, and student 1 ranks more schools than student 2 because student 1 is more eager and higher achieving (regardless of whether she attends A). Whenever $First_A(r) = \{1, 2\}$, student 1 gets the seat at A and student 2 does not. Comparing 1 and 2 within $First_A(r) = \{1, 2\}$, the researcher is likely to mistakenly conclude A has a positive achievement effect.

Such a correlation between preferences and outcomes is empirically observed in data from Denver Public Schools. Denver Public Schools use the DA mechanism for unified public and charter school admissions (Abdulkadiroğlu et al., 2017). Each year more than 10000 applicants in grades 4-10 participate in this system. These applicants are predominantly black and hispanic and from needy households. I use Denver's data for school years 2011-2013 to correlate applicant preference lengths and pre-application baseline test scores, which are likely to be predictive of potential outcomes after mechanism participation.

There turns out to be a clear correlation between preference length and baseline scores, as seen in Table 2. For all of math, reading, and writing, students with higher baseline scores tend to rank more schools; there is about 0.15 standard deviation score difference between students who rank only one school and those who rank two or more schools. This is reasonable if, for example, higher-achieving students are more willing to investigate and rank schools because of smaller learning costs. This empirically suggests that student type is correlated with potential outcomes and is a source of potential omitted variable bias, as the above theoretical story assumes.

Table 2: Empirical Correlation between Preferences and Outcomes

# of Schools Ranked	Average Baseline Test Scores (Standardized)			# of Students
	Math	Reading	Writing	
1	-0.10	-0.09	-0.11	8051
≧2	0.05	0.04	0.04	14497

Notes: This table shows average baseline test scores for students who rank different numbers of schools. Each test score is standardized to the test score distribution for the whole population of students in Denver Public Schools.

The above example raises the question: Under what circumstances does the first-choice research design extract a random assignment as desired?

3.2 Strategy-proofness for Schools

The success or failure of the first-choice research design turns out to be linked to a seemingly unrelated property of mechanisms. So far, I have treated priorities and lottery numbers as public information. In this section, I depart from this assumption and imagine a hypothetical situation in which schools have priorities and lottery numbers as their private information. The priorities and lottery numbers are assumed to represent school preferences; I come back to the interpretation of this thought experiment at the end of this section. Suppose a gDA mechanism asks schools to report priorities and lottery numbers. Their reports are not necessarily truthful. The gDA mechanism then uses the reported priorities and lottery numbers to create a matching.

Given any $(I, S, \succ_I, (c_s)_{s\in S})$, let $\Gamma \equiv \{(\rho_s, r_s) \in \{1, ..., K\}^{|I|} \times [0,1]^{|I|} | \rho_{is} + r_{is} \neq \rho_{js} + r_{js}$ for all students $i \neq j\}$ be the domain of possible priority and lottery number reports. This domain specification implies every student is acceptable to every school in any reported priority and lottery numbers. A gDA mechanism asks each school s to report its priority and lottery numbers (ρ_s, r_s), producing $(\rho, r) \equiv (\rho_s, r_s)_{s\in S}$. Let $\varphi(\rho, r) \equiv (\varphi_s(\rho, r))_{s\in S}$ be the assignment produced by a gDA mechanism φ for the reported priority and lottery numbers (ρ, r).

School s's preference $\succ_s$, which is defined over *the set of subsets of* I, is said to be responsive with respect to (c_s, ρ_s, r_s) (Roth and Sotomayor, 1992) if the following holds.

(1) For any $i, i' \in I$, if $\rho_{is} + r_{is} < \rho_{i's} + r_{i's}$, then for any $I' \subseteq I \setminus \{i, i'\}$, $I' \cup \{i\} \succ_s I' \cup \{i'\}$,

(2) $\emptyset \succ_s I'$ for any $I' \subseteq I$ with $|I'| > c_s$, and

(3) For any $I' \subseteq I$ with $|I'| < c_s$ and any $i \in I \setminus I'$, it holds $I' \cup \{i\} \succ_s I'$.

I use these concepts to define the following property.

Definition 3. A gDA mechanism φ is **strategy-proof for school** s if for any $(I, S, \succ_I, (c_s)_{s \in S})$, any priority and lottery number profile $(\rho^*, r^*) \in \Gamma^{|S|}$, any preference $\succ_s^*$ responsive with respect to (c_s, ρ_s^*, r_s^*), and any $(\rho_s', r_s') \in \Gamma$,

$$\varphi_s(\rho^*, r^*) \succeq_s^* \varphi_s((\rho_s', r_s'), (\rho_{-s}^*, r_{-s}^*)),$$

where $\succeq_s^*$ is the weak preference associated with $\succ_s^*$ and $(\rho_{-s}^*, r_{-s}^*) \equiv (\rho_{s'}^*, r_{s'}^*)_{s' \neq s}$. A gDA mechanism φ is strategy-proof for schools if it is strategy-proof for every school s.

This definition of strategy-proofness is a non-stochastic, ex post property though my setting has stochastic elements due to lotteries. The standard behavioral interpretation of this concept is that no school ever has a preference manipulation that is profitable with respect to its true preference. It is crucial to note, however, that I am *not* concerned with this usual interpretation. As will become clearer in the next section, unlike usual studies on strategy-proofness, I am interested only in the mathematical implications of strategy-proofness for empirical research. These implications are true regardless of whether strategy-proofness itself has any relevance as an incentive compatibility property or desideratum. As a result, the following usual questions about strategy-proofness for schools are all irrelevant for my analysis: Do schools have preferences? Are school preferences consistent with priorities? Do schools ever game the system?

3.3 Sufficiency: Strategy-proofness Generates Natural Experiments

Strategy-proofness for schools turns out to be sufficient for the first-choice research design to extract a random assignment.

Theorem 1. *The first-choice research design extracts a random assignment for a gDA mechanism φ if φ is strategy-proof for schools.*

The proof is in Appendix A.1. Combined with existing results on strategy-proofness for schools, Theorem 1 provides positive results for the first-choice research design for some of the gDA mechanisms.

Corollary 1. *a) The first-choice research design extracts a random assignment for the Boston mechanism with any lottery regime.*
b) The first-choice research design extracts a random assignment for the DA mechanism with STB when there are no priorities ($\rho_{is} = \rho_{js}$ for all students i, j and school s). This mechanism is often called random serial dictatorship.

Proof. (a) follows from Theorem 1 and Ergin and Sönmez (2006)'s Theorem 2 that the Boston mechanism is strategy-proof for schools. (b) follows from the proof of Theorem 1 and the fact that for the DA mechanism, truth-telling is optimal for any school s when all the other schools report the same preference as s's true preference. See Appendix A.2 for details. □

I illustrate Theorem 1 with the Boston mechanism. Consider Example 1 in Section 3.1 and a thought experiment where schools have private preferences and the mechanism asks schools to report their preferences. First of all, school A is never matched with student 3 since 3 ranks A second and the seat at A is always filled by one of the two students who rank A first. A is thus matched with either 1 or 2. When A's true preference is such that $1 \succ_A 2$, A is matched with the more preferred student 1 by truth-telling.[16] When A's true preference is with $2 \succ_A 1$, A is matched with the more preferred student 2 by truth-telling. Therefore, there is no profitable preference manipulation for A; the Boston mechanism is strategy-proof for A in Example 1 (Ergin and Sönmez, 2006).

As it should be by strategy-proofness and Theorem 1, the first-choice research design extracts a random assignment for A in Example 1 for the Boston mechanism. Note that $First_A(r) = \{1, 2\}$ for all r since only 1 and 2 rank A first with the same priority and only one of them with a better lottery number is assigned A under any r. Enumerating all lottery outcomes shows that 1 and 2 share the same assignment probability of 1/2 at A, i.e., $P(D_{1A}(R) = 1) = P(D_{2A}(R) = 1) = 1/2$. Therefore the first-choice research design extracts a random assignment.

Illustrative Proof of a Special Case of Theorem 1

Readers who are not interested in the proof may skip the remainder of this section and jump to Section 3.4. The full proof of Theorem 1 is long and involved. For purposes of illustration, this section provides a simpler proof for a special case of Theorem 1. The special case of interest is formulated as follows.

[16] Since A's capacity is 1, I do not need to distinguish its preference over sets of students and its priority order over individual students.

Corollary 2. *Consider assignment problems with no priorities ($\rho_{is} = \rho_{js}$ for all students i and j and school s) and unit school capacities ($c_s = 1$ for all s). The first-choice research design extracts a random assignment for a gDA mechanism φ with a multiple tie breaker (MTB) if φ is strategy-proof for schools.*

A formal proof of this fact needs a few definitions. Given any gDA mechanism φ and lottery number profile r, I say two students i_0 and i_1 are **consecutive in** r_s **within** $First_s(r)$ if $i_0, i_1 \in First_s(r)$ and there is no other student $j \in First_s(r)$ such that $r_{i_0s} < r_{js} < r_{i_1s}$ or $r_{i_0s} > r_{js} > r_{i_1s}$. As per usual, a **permutation** of r_s is a bijection from $\{r_{is}\}_{i\in I}$ to $\{r_{is}\}_{i\in I}$ itself. A permutation r'_s of r_s is said to be a **first-choice transposition** of r_s at r if r'_s switches only two students i_0 and i_1 who are consecutive in r_s within $First_s(r)$, i.e., $r'_{i_0s} = r_{i_1s}, r'_{i_1s} = r_{i_0s}$, and $r'_{js} = r_{js}$ for all $j \neq i_0, i_1$. I use these definitions to introduce a key property of mechanisms.

Definition 4. Suppose that the assumptions in Corollary 2 hold. I say a gDA mechanism φ with MTB **satisfies the Fisher property** if given φ, for any $(I, S, \succ_I)$, any school s, any lottery number profile r and any first-choice transposition r'_s of r_s at r that switches only two students i_0 and i_1, the following is true:

- $D_{i_1s}(r'_s, r_{-s}) = D_{i_0s}(r)$,
- $D_{i_0s}(r'_s, r_{-s}) = D_{i_1s}(r)$, and
- $D_{js}(r'_s, r_{-s}) = D_{js}(r)$ for all $j \neq i_0, i_1$.

In words, a gDA mechanism with MTB satisfies the Fisher property if for that mechanism, any transposition of lottery numbers within the first-choice subsample translates into the same transposition of assignments. Since any permutation is a combination of transpositions, the Fisher property implies that any permutation of lottery numbers in the first-choice subsample always induces the same permutation of assignments. I name this the Fisher property after Ronald Fisher (the inventor of randomized experiments) since this property is reminiscent of a randomized controlled trial, where random numbers pin down treatment assignment. Note that the Fisher property implies that for any $(I, S, \succ_I)$, any school s, any lottery number profile r and any first-choice transposition r'_s of r_s at r,

$$First_s(r) = First_s(r'_s, r_{-s}). \tag{1}$$

Not surprisingly, I find that if a gDA mechanism with MTB satisfies the Fisher property, then the first-choice research design extracts a random assignment for that mechanism.

Moreover, less trivially, strategy-proofness for schools turns out to imply the Fisher property, showing Corollary 2. I prove these facts in the following proof of Corollary 2.

Proof of Corollary 2. The proof consists of two steps, Lemmas 1 and 2 below.

Lemma 1. *Under the assumptions in Corollary 2, the first-choice research design extracts a random assignment for a gDA mechanism φ with MTB if φ satisfies the Fisher property.*

Proof of Lemma 1. Consider any assignment problem X and let $\mathcal{R} \equiv \{r \in [0,1]^{|I|\times|S|} | r_{is} \neq r_{js}$ for all students i, j, and school $s\}$ be the set of all possible values of the lottery number profile r. Fix any gDA mechanism φ with MTB and the Fisher property, any school s, and any potential lottery realization r. Partition $\mathcal{R}$ into $\mathcal{P} \equiv \{\mathcal{R}_n\}_{n\in\mathcal{N}}$ ($\mathcal{N}$ is an uncountable set of indices) such that the following holds: Within each $\mathcal{R}_n$, any $r' \in \mathcal{R}_n$ can be obtained from any other $r'' \in \mathcal{R}_n$ by applying a finite number of first-choice transpositions at school s, i.e., there exists a sequence of lottery number profiles $(r^1, r^2, ..., r^K)$ such that $r^1 = r'', r^K = r'$, and for each $k = 2, ..., K$, r^k_s is a first-choice transposition of r^{k-1}_s at r^{k-1}.[17]

The Fisher property and equation (1) imply that conditional on each $\mathcal{R}_n$, $First_s(r_n)$ and $o_s(r_n) \equiv |\{i \in First_s(r_n)|D_{is}(r_n) = 1\}|$ are constant for all $r_n \in \mathcal{R}_n$. This means that for each $r_n \in \mathcal{R}_n$, students with the $o_s(r_n)$-best lottery numbers in $First_s(r_n)$ have $D_{is}(r_n) = 1$, which happens with probability $\dfrac{o_s(r_n)}{|First_s(r_n)|}$ for any $i \in First_s(r_n)$ conditional on $\mathcal{R}_n$. Also, whenever $First_s(r_n)$ and $First_s(r)$ are nonempty, $First_s(r) = First_s(r_n) = \{i|rank_{is} = 1\}$ by the no-priority assumption. Therefore, for each $n \in \mathcal{N}$,

$$
\begin{aligned}
&P(D_{is}(R) = 1|i \in First_s(r), R \in \mathcal{R}_n, \theta_i = \theta) \\
&= \begin{cases} \dfrac{o_s(r_n)}{|First_s(r_n)|} & \text{if } First_s(r_n) \neq \emptyset \\ 1 & \text{if } First_s(r_n) = \emptyset \text{ and } D_{js}(r_n) = 1 \text{ for all } r_n \text{ and all } j \in First_s(r) \\ 0 & \text{if } First_s(r_n) = \emptyset \text{ and } D_{js}(r_n) = 0 \text{ for all } r_n \text{ and all } j \in First_s(r) \end{cases} \\
&\equiv p_n,
\end{aligned}
$$

which is independent of θ_i.[18]

[17] This partition is well-defined since the assumption that φ satisfies the Fisher property guarantees equation (1), which in turn implies that r' can be obtained from r'' with a finite number of first-choice transpositions if and only if r'' can be obtained from r' with a finite number of first-choice transpositions. Note also that this partition depends on particular school s I focus on.

[18] In the above definition of p_n, the three cases are exhaustive by the following reason. Whenever $First_s(r_n) = \emptyset$, it has to be the case that $\mathcal{R}_n = \{r_n\}$ (i.e., $\mathcal{R}_n$ is a singleton) since there is no first-choice transposition of r_{ns} with $First_s(r_n) = \emptyset$. For the single element r_n, there are only two possibilities, (i) $D_{js}(r_n) = 1$ for all $j \in First_s(r)$ or (ii) $D_{js}(r_n) = 0$ for all $j \in First_s(r)$. To see this, suppose to the

Let Y be $Y(R) = \mathcal{R}_n^R$ where $\mathcal{R}_n^R$ is the element of the partition $\mathcal{P}$ with $R \in \mathcal{R}_n^R$. Let P_Y be the probability measure of Y induced by that of R, i.e. for all $A \subset \mathcal{N}$, $P_Y(\{\mathcal{R}_n\}_{n\in A}) \equiv P(R \in \cup_{n\in A}\mathcal{R}_n)$. With this notation, I have

$$\begin{aligned}
&P(D_{is}(R) = 1 | i \in First_s(r), \theta_i = \theta)\\
&= \int_{\{\mathcal{R}_n\}_{n\in\mathcal{N}}} P(D_{is}(R) = 1 | i \in First_s(r), Y = \mathcal{R}_n, \theta_i = \theta) dP_Y(\mathcal{R}_n)\\
&\quad \text{(by the law of iterated expectation)}\\
&= \int_{\{\mathcal{R}_n\}_{n\in\mathcal{N}}} p_n dP_Y(\mathcal{R}_n),
\end{aligned}$$

which is again independent of θ_i since both p_n and P_Y are independent of θ_i. Thus $P(D_{is}(R) = 1 | i \in First_s(r), \theta_i = \theta) = P(D_{is}(R) = 1 | i \in First_s(r))$, and the first-choice research design extracts a random assignment. This completes the proof of Lemma 1. $\square$

Lemma 2. *Under the assumptions in Corollary 2, a gDA mechanism φ with MTB satisfies the Fisher property if φ is strategy-proof for schools.*

Proof of Lemma 2. As in Definition 4 of the Fisher property, consider any gDA mechanism φ with MTB, any $(I, S, \succ_I)$, any school s, any lottery number profile r, and any first-choice transposition r'_s of r_s at r that switches only two students i_0 and i_1. Assume that φ is strategy-proof for schools. Without loss of generality, assume $r_{i_1 s} < r_{i_0 s}$, i.e, student i_1 has a better original lottery number than i_0 at school s. This assumption makes it impossible that $D_{i_0 s}(r) = 1$ and $D_{i_1 s}(r) = 0$ since $i_0, i_1 \in First_s(r)$ and so $\rho^\varphi_{i_0 s} = \rho^\varphi_{i_1 s}$ and both i_0 and i_1 rank s first. $D_{i_0 s}(r) = D_{i_1 s}(r) = 1$ is also impossible by the unit capacity assumption of $c_s = 1$. There are two remaining cases, Cases i and ii below.

<u>*Case i*</u>: $D_{i_0 s}(r) = 0$ and $D_{i_1 s}(r) = 1$, i.e., i_1 is assigned to s while i_0 is not under r. To reach the desired Fisher property, consider the DA algorithm inside the gDA mechanism with (r'_s, r_{-s}). In the first round of the DA algorithm, students i_0 and i_1 apply to school s (as it is their first choice by $i_0, i_1 \in First_s(r)$). School s's single seat is tentatively assigned to i_0; it is because (a) $D_{i_1 s}(r) = 1$ and so student i_1 is tentatively assigned to school s in the first round of the DA algorithm under r, (b) the same set of applicants apply to school s in the first round of the DA algorithm with (r'_s, r_{-s}) and r, and (c) $\rho^\varphi_{i_0 s} + r'_{i_0 s} = \rho^\varphi_{i_1 s} + r_{i_1 s} < \rho^\varphi_{i_1 s} + r'_{i_1 s} = \rho^\varphi_{i_0 s} + r_{i_0 s}$ while $\rho^\varphi_{js} + r'_{js} = \rho^\varphi_{js} + r_{js}$ for all $j \neq i_0, i_1$. The tentative assignment of i_0 to s will never be canceled, as I claim below.

contrary that there are $j, k \in First_s(r)$ such that $D_{js}(r_n) \neq D_{ks}(r_n)$. Then j and k must be in $First_s(r_n)$ by the definition of the first-choice subsample. But this is a contradiction to $First_s(r_n) = \emptyset$.

Claim 1. $D_{i_0 s}(r'_s, r_{-s}) = 1$, i.e., school s's single seat is assigned to i_0 at the end of the DA algorithm with (r'_s, r_{-s}).

Proof of Claim 1. Suppose not. In second or later rounds of the DA algorithm with (r'_s, r_{-s}), school s must reject i_0 in favor of some student $j(\neq i_0, i_1)$ with better lottery number $r'_{js} < r'_{i_0 s}$. This implies by the construction of first-choice transposition r'_s that $r_{js} = r'_{js} < r'_{i_0 s} = r_{i_1 s}$. In other words,

$$\{j\} \succ_s \{i_1\} \tag{2}$$

for any preference $\succ_s$ responsive with respect to (c_s, ρ_s, r_s) where $c_s = 1$ and $\rho_{is} = \rho_{ks}$ for all students i and k as required by the assumptions in Corollary 2.

The definition of the DA algorithm with (r'_s, r_{-s}) ensures that the tentatively assigned student at s improves with respect to r'_s and so with respect to any preference $\succ'_s$ responsive with respect to (c_s, ρ_s, r'_s). For any such preference $\succ'_s$, therefore, it is the case that $\varphi_s(r'_s, r_{-s}) \succeq'_s \{j\}$. This implies that for any preference $\succ_s$ responsive with respect to (c_s, ρ_s, r_s),

$$\varphi_s(r'_s, r_{-s}) \succeq_s \{j\}. \tag{3}$$

This is because (a) $k \neq i_0, i_1$ for student k defined by $\{k\} = \varphi_s(r'_s, r_{-s})$ and (b) by the construction of r'_s, for any students $j, h \neq i_0, i_1$, I have $r_{hs} < r_{js}$ if and only if $r'_{hs} < r'_{js}$. Combining these steps (2), (3), and $D_{i_1 s}(r) = 1$ together, for any preference $\succ_s$ responsive with respect to (c_s, ρ_s, r_s), I have

$$\varphi_s(r'_s, r_{-s}) \succeq_s \{j\} \succ_s \{i_1\} = \varphi_s(r). \tag{4}$$

However, the preference relation (4) contradicts the assumption that φ is strategy-proof for schools. Therefore, student i_0 will be assigned to school s, proving Claim 1. □

This Claim, $c_s = 1$, and the assumption of $D_{i_0 s}(r) = 0$ and $D_{i_1 s}(r) = 1$ jointly imply that

- $D_{i_1 s}(r'_s, r_{-s}) = D_{i_0 s}(r) = 0$,
- $D_{i_0 s}(r'_s, r_{-s}) = D_{i_1 s}(r) = 1$, and
- $D_{js}(r'_s, r_{-s}) = D_{js}(r) = 0$ for all $j \neq i_0, i_1$,

proving the Fisher property for Case i.

Case ii: $D_{i_0s}(r) = D_{i_1s}(r) = 0$, i.e., neither i_0 nor i_1 is assigned to s at r. This means that for any preference $\succ_s$ responsive with respect to (c_s, ρ_s, r_s) (with $c_s = 1$ and $\rho_{is} = \rho_{ks}$ for all students i and k), I have

$$\varphi_s(r) \succ_s \{i_0\} \text{ and } \varphi_s(r) \succ_s \{i_1\}, \tag{5}$$

since school s tentatively keeps a student while rejecting i_0 and i_1, both of whom are in $First_s(r)$ and apply for s at the first step of the DA algorithm with r. To show the Fisher property, suppose to the contrary that the Fisher property does not hold, i.e., $D_{i_1s}(r'_s, r_{-s}) \neq D_{i_0s}(r)$ or $D_{i_0s}(r'_s, r_{-s}) \neq D_{i_1s}(r)$ or $D_{js}(r'_s, r_{-s}) \neq D_{js}(r)$ for some $j \neq i_0, i_1$. There are two sub-cases to discuss.

Case ii.a: $D_{i_1s}(r'_s, r_{-s}) \neq D_{i_0s}(r) = 0$ or $D_{i_0s}(r'_s, r_{-s}) \neq D_{i_1s}(r) = 0$. This requires $D_{i_0s}(r'_s, r_{-s}) = 1$ and $D_{i_1s}(r'_s, r_{-s}) = 0$ by the assumption of $r'_{i_0s} = r_{i_1s} < r_{i_0s} = r'_{i_1s}$ and the unit capacity assumption of $c_s = 1$. Equivalently, $\varphi_s(r'_s, r_{-s}) = \{i_0\}$. Together with the preference relation (5), it has to be the case that for any preference $\succ'_s$ responsive with respect to (c_s, ρ_s, r'_s), I have

$$(\{k\} \equiv)\varphi_s(r) \succ'_s \{i_0\} = \varphi_s(r'_s, r_{-s}). \tag{6}$$

or, equivalently, $r'_{ks} < r'_{i_0s}$. This is because student k has $r_{ks} < r_{i_0s}, r_{i_1s}$ by the preference relation (5) and the construction of r'_s guarantees $r'_{ks} < r'_{i_0s}, r'_{i_1s}$. However, the preference relation (6) contradicts the assumption that φ is strategy-proof for schools. Therefore, the Fisher property must hold for Case ii.a.

Case ii.b: $D_{i_1s}(r'_s, r_{-s}) = D_{i_0s}(r) = 0$ and $D_{i_0s}(r'_s, r_{-s}) = D_{i_1s}(r) = 0$, but $D_{js}(r'_s, r_{-s}) \neq D_{js}(r)$ for some $j \neq i_0, i_1$. This implies $\varphi_s(r'_s, r_{-s}) \neq \varphi_s(r)$ while $i_0, i_1 \notin \varphi_s(r'_s, r_{-s}) \cup \varphi_s(r)$. The construction of r'_s implies that for any preferences $\succ_s$ and $\succ'_s$ responsive with respect to (c_s, ρ_s, r_s) and (c_s, ρ_s, r'_s), respectively, I have

$$[\varphi_s(r'_s, r_{-s}) \succ_s \varphi_s(r) \text{ and } \varphi_s(r'_s, r_{-s}) \succ'_s \varphi_s(r)] \tag{7}$$

or

$$[\varphi_s(r) \succ_s \varphi_s(r'_s, r_{-s}) \text{ and } \varphi_s(r) \succ'_s \varphi_s(r'_s, r_{-s})]. \tag{8}$$

Both (7) and (8) contradict the assumption that φ is strategy-proof for schools. Thus the Fisher property must hold, completing the proof of Lemma 2. □

Lemmas 1 and 2 jointly prove Corollary 2. □

This proof of Corollary 2 is still far from a complete proof of Theorem 1, however. For instance, Theorem 1 allows additional complications like priorities, non-unit capacities, and STB, all of which the above proof ignores. Nevertheless, the proof in Appendix A.1 shows that the sufficiency of strategy-proofness generally holds.

3.4 Necessity in Practice

Theorem 1 shows that strategy-proofness for schools is sufficient for the first-choice research design to extract a random assignment. Strategy-proofness turns out to be not only sufficient but also necessary as long as I focus on practically important mechanisms.

Proposition 1. *Even with unit school capacities ($c_s = 1$ for all s), the first-choice research design does not extract a random assignment for the DA, Charlotte, and "top trading cycles" mechanisms (with any lottery regime), all of which are known to be not strategy-proof for schools.*

To illustrate this, consider the DA mechanism in Example 1. Imagine A's true preference is $3 \succ_A 1 \succ_A 2$ while B's is $1 \succ_B 2 \succ_B 3$. Under these true preferences, A is matched with 1. If A misreports $3 \succ'_A 2 \succ'_A 1$, however, A is matched with 3, the most preferred student with respect to $\succ_A$. Therefore, the DA mechanism is not strategy-proof for A in Example 1. This confirms the classic result that the DA mechanism is not strategy-proof for schools (Roth and Sotomayor, 1992).

Intuitively, school A benefits from manipulating its preference and rejecting 1 by the following chain reaction of rejections and new applications. After being rejected by A, student 1 next applies for B, which results in B's rejecting 3. Student 3, the most preferred student for A, then applies for and benefits A.

The same chain reaction causes the first-choice research design to fail, as I explain in Section 3.1. Different applicants cause different chain reactions that have different effects on assignment probabilities at A, depending on schools ranked below A. This can cause applicants in $First_A(r_0)$ to have different assignment probabilities at A.[19]

[19]In contrast, for the Boston mechanism analyzed in the last section, such chain reactions do not affect assignments to A. By its construction, for the Boston mechanism, each school is forced to prioritize students ranking it higher over students ranking it lower. As a result, chain reactions caused by student i at schools ranked below A involve only students who rank A lower than student i does. Any such student in chain reactions i causes is never accepted by A since A rejects i. Thus, different chain reactions caused by different students have no effect on assignments at A. This is the reason why the Boston mechanism is strategy-proof for schools and the first-choice research design extracts a random assignment for the Boston mechanism.

Similarly, the first-choice research design does not extract a random assignment for the Charlotte mechanism, another mechanism that is not strategy-proof for schools. Suppose that in Example 1, student 3's priority at A and 1's priority at B are walk zone priorities. In this case, the Charlotte mechanism coincides with the DA mechanism. The Charlotte mechanism is therefore manipulable by schools and the first-choice research design fails in the same way as for the DA mechanism. Section 5.3 also shows the same failure of the first-choice design for the top trading cycles mechanism, which is not strategy-proof for schools either. In these senses, for mechanisms frequently discussed in theory and practice, strategy-proofness for schools is required for the first-choice research design to extract a random assignment.[20] The analyst therefore needs to take care when using the first-choice design for non-strategy-proof mechanisms such as the DA, Charlotte, and top trading cycles mechanisms.

Empirical Illustration

The DA mechanism is not strategy-proof for schools and may not extract a random assignment via the first-choice design. To see whether the theoretical result has any relevance in practice, I use data from Denver Public Schools, which use the usual DA mechanism with STB for unified public and charter school admissions (Abdulkadiroğlu et al., 2017). I use its DA mechanism in school years 2011-2012 as follows.

(1) Taking student preferences, school priorities, and capacities as fixed, I simulate the DA mechanism by drawing counterfactual lottery numbers one million times. This gives me an approximate assignment probability $\hat{P}(D_{is}(R) = 1)$ for each student i and school s, i.e., the empirical frequency of student i's being assigned to s over the one million simulations.[21]

(2) Let r_0 be the realized lottery number profile Denver Public Schools drew for the year. For each school s and each student i in the realized first-choice subsample $First_s(r_0)$ (if any), I demean i's assignment probability by subtracting the mean of assignment probabilities at s across all students in $First_s(r_0)$. That is, I compute

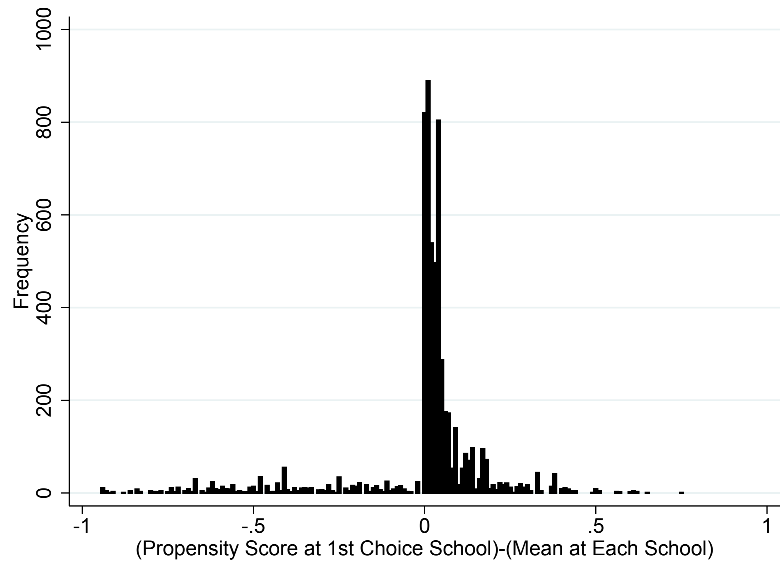
$$\hat{P}^{demean}(D_{is}(R) = 1) = \hat{P}(D_{is}(R) = 1) - \frac{\Sigma_{j \in First_s(r_0)} \hat{P}(D_{js}(R) = 1)}{|First_s(r_0)|}.$$

[20] On the other hand, strategy-proofness for schools is not exactly necessary. See Appendix B.2 for details.

[21] In Denver, each school is divided into multiple sub-schools (called "buckets") with their own priorities and capacities. Buckets correspond to schools in my theoretical model. Below I use "schools" to mean buckets. See Abdulkadiroğlu et al. (2017) for more details of the Denver school admissions system.

(3) I plot this assignment probability deviation $\hat{P}^{demean}(D_{is}(R) = 1)$ across all schools s and all students in $First_s(r_0)$.

Figure 1: Empirical Illustration of Proposition 1



The resulting histogram of assignment probability deviations is in Figure 1. If the 1st choice method extracts a random assignment, $\hat{P}(D_{is}(R) = 1) \approx \hat{P}(D_{js}(R) = 1)$ for all s and all $i, j \in First_s(r_0)$ and so the assignment probability deviation $\hat{P}^{demean}(D_{is}(R) = 1)$ would be almost 0 (up to simulation errors) for all s and all i in $First_s(r_0)$. As the figure shows, however, there are many values of $\hat{P}^{demean}(D_{is}(R) = 1)$ that are far from 0. The standard deviation is around 0.19. This provides an empirical illustration of the theoretical necessity of strategy-proofness for schools.

4 Qualification Instrumental Variable Research Design

While the previous sections focus on the first-choice research design, several empirical studies use an alternative research design. I call this alternative the qualification instrumental variable (IV) research design.[22] Unlike the first-choice research design (which tries to make assignments random by focusing on a subset of students), the qualification IV research design considers all students and tries to code a random instrumental variable for non-random assignments. Define the qualification IV by

[22]For empirical examples of the qualification IV design, see Pop-Eleches and Urquiola (2013); Dobbie and Fryer (2014); Lucas and Mbiti (2014), all of which use data from the DA mechanism.

$$Z_{is}(r) \equiv 1\{\rho^{\varphi}_{is} + r_{is} \leq \max\{\rho^{\varphi}_{js} + r_{js} | D_{js}(r) = 1\}\}.$$

If there is no student j with $D_{js}(r) = 1$, then define $Z_{is}(r) = 1$ for all i. The qualification IV for a student at a school is turned on if her realized priority rank at the school is better than that of some student assigned to the school. Note that $Z_{is}(r) = 1$ is possible even for students who do not apply to school s.

The qualification IV looks unconfounded conditional on ρ^{φ}_{is}. The IV is also likely to be correlated with assignment $D_{is}(r)$ since i can get assigned only when she is qualified ($Z_{is}(r) = 1$). Based on this idea, the qualification IV research design instruments for assignment $D_{is}(r)$ by qualification $Z_{is}(r)$ conditional on ρ^{φ}_{is}. The design then estimates treatment effects by Two Stage Least Square or other instrumental variable models (Heckman and Vytlacil (2007) chapter 4, Manski (2008) chapter 3, Angrist and Pischke (2009) chapter 4).[23] For this research design to identify a causal effect, the qualification IV needs to be random conditional on ρ^{φ}_{is}, as formalized in the following definition.

Definition 5. The qualification IV research design extracts a random assignment for a gDA mechanism φ for school s at assignment problem X if given φ and X, for all modified priority ρ and student type θ,

$$P(Z_{is}(R) = 1 | \rho^{\varphi}_{is} = \rho, \theta_i = \theta) = P(Z_{is}(R) = 1 | \rho^{\varphi}_{is} = \rho).$$

An equivalent requirement is

$$P(Z_{js}(R) = 1) = P(Z_{ks}(R) = 1)$$

for all students $j, k \in I$ with $\rho^{\varphi}_{js} = \rho^{\varphi}_{ks}$. The qualification IV research design extracts a random assignment for a gDA mechanism φ if it does so for every school s at every problem X.

This property requires that conditional on modified priority status ρ^{φ}_{is}, qualification for school s is random and independent of students' preferences and priorities contained in θ_i.

[23] That is, for outcome Y_i of interest and the realized lottery outcome r_0 in the data, the qualification IV research design uses the following Two Stage Least Square regression or a similar IV model:

$Y_i = \alpha_2 + \beta_2 D_{is}(r_0) + \Sigma_k \gamma^k_2 1\{\rho^{\varphi}_{is} = k\} + \epsilon_{2i}$ (second stage regression)
$D_{is}(r_0) = \alpha_1 + \beta_1 Z_{is}(r_0) + \Sigma_k \gamma^k_1 1\{\rho^{\varphi}_{is} = k\} + \epsilon_{1i}$ (first stage regression)

Only under this conditionally random assignment does the qualification IV Z_{is} generate exogenous or random variation in assignment D_{is}.[24] It turns out that no gDA mechanism satisfies the above property even in the simple case with no priorities and unit capacities.

Proposition 2. *Consider any sets of at least three students and at least three schools. Even with no priorities ($\rho_{is} = \rho_{js}$ for all students i, j, and school s) and unit school capacities ($c_s = 1$ for all s), there exist student preference profiles at which every student ranks some schools and the following holds: There is no gDA mechanism with any lottery regime for which the qualification IV research design extracts a random assignment.*

The proof is in Appendix A.3. I require every student to rank some school for excluding uninteresting cases where students rank no school and have no effect on assignment outcomes. I illustrate this result by an example.

Example 2. There are applicants 1, 2, 3, 4, 5 and schools A, B, C with the following preferences and priorities:

$$\begin{aligned}
&\succ_1 : B, A, \emptyset \\
&\succ_2 : B, \emptyset \\
&\succ_3 : C, A, \emptyset \\
&\succ_4, \succ_5 : C, \emptyset \\
&\rho_A, \rho_B, \rho_C : \{1, 2, 3, 4, 5\}.
\end{aligned}$$

The capacity of each school is 1. The treatment school is A.

Students 1 and 3 share the same modified priority ρ^{φ}_{iA} for any gDA mechanism φ: Both students 1 and 3 rank A second and have the same priority at A so that $\rho^{\varphi}_{1A} \equiv f^{\varphi}(\rho_{1A}) + g^{\varphi}(rank_{1A}) = f^{\varphi}(\rho_{3A}) + g^{\varphi}(rank_{3A}) \equiv \rho^{\varphi}_{3A}$, which I denote by ρ. Nevertheless, enumerating all possible lottery orders shows that for any gDA mechanism, we have

$$P(Z_{iA}(R) = 1 | \rho^{\varphi}_{iA} = \rho, \theta_i = \theta_1) = 2/3 \neq 5/6 = P(Z_{iA}(R) = 1 | \rho^{\varphi}_{iA} = \rho, \theta_i = \theta_3) \text{ for STB}$$
$$P(Z_{iA}(R) = 1 | \rho^{\varphi}_{iA} = \rho, \theta_i = \theta_1) = 2/3 \neq 3/4 = P(Z_{iA}(R) = 1 | \rho^{\varphi}_{iA} = \rho, \theta_i = \theta_3) \text{ for MTB.}$$

[24] Definition 5 for the qualification IV design may appear to be incomparable with Definition 2 for the first-choice design. However, Appendix B.1 shows that these two definitions are special cases of a unified definition of a random assignment under general empirical research designs, including the first-choice and qualification IV designs. Hence, it is legitimate to use Definitions 2 and 5 to compare the two research designs.

A computer program to implement this computation is available upon request. Therefore, even with no priorities and unit capacities, the qualification IV research design does not extract a random assignment for any gDA mechanism.[25]

Intuitively, the qualification IV research design fails in this example because students 1 and 3 experience different levels of competition at their first-choice schools B and C, respectively, before applying for A. Let me consider the following cases.

Case i: Neither student 1 nor 3 applies for A, i.e., 1 and 3 are assigned B and C, respectively. In this case, no student applies for A, and A is undersubscribed. Both 1 and 3 are therefore qualified for A.

Case ii: Only student 1 applies for A. In this case, 1 is always assigned A and qualified for A. By $\rho^{\varphi}_{1A} = \rho^{\varphi}_{3A}$ shown above, student 3 is qualified for A if and only if 3 has a better lottery number than 1 at A.

Case iii: Only student 3 applies for A. In this case, 3 is always assigned A and qualified for A. Student 1 is qualified for A if and only if 1 has a better lottery number than 3 at A.

Case iv: Both students 1 and 3 apply for A. In this case, only one of 1 and 3 with a better lottery number is assigned A and qualified for A.

For simplicity, consider the MTB lottery regime. Cases i and iv are ignorable since they do not cause any difference between 1's and 3's qualification probabilities at A. Conditional on Case ii, student 1 is qualified for sure while 3 is qualified with probability 0.5, the probability that 3 has a better lottery number than 1 for A. Likewise, conditional on Case iii, student 3 is qualified for sure, but 1 is qualified only with probability 0.5. Crucially, Case iii is more likely to happen than Case ii. This is because 3's first choice (C) is more competitive than 1's first choice (B) and so 3 is more easily rejected by the first choice and more likely to apply for the second-choice school, A. As a result, 3 is more likely to be qualified for A than 1 due to differential competition at their first-choice schools, as the proof in Appendix A.3 makes it precise. The proof also generalizes this observation

[25] Since both $\rho^{\varphi}_{1A} = \rho^{\varphi}_{3A}$ and $\rho_{1A} = \rho_{3A}$ hold, the counterexample works even if I use original priorities to define an alternative qualification IV as $Z'_{is}(r) \equiv 1\{\rho_{is} + r_{is} \leq \max\{\rho_{js} + r_{js} | D_{js}(r) = 1\}\}$. Also, note that students 1 and 3 share the same priority at all schools in the above example. Thus, the qualification IV research design may fail even if I modify it to the more refined version that conditions on having the same priority at all schools. Finally, the qualification IV research design does not extract a random assignment even for the top trading cycles mechanism, as shown in Section 5.3.

to any lottery regime and any market size. Section 5.2 shows that the problem with the qualification IV holds up even if I modify its definition, e.g., by changing the priority cutoff $\max\{\rho^{\varphi}_{js} + r_{js} | D_{js}(r) = 1\}$ to a constant number.

Example 2 illustrates a general point that students may have different qualification probabilities depending on which schools they rank higher than the treatment school. This does not matter for the first-choice design since students in the first-choice subsample $First_s(r_0)$ rank no school above the treatment school. Therefore, the above trouble does not happen to the first-choice research design focusing on the first-choice subsample $First_s(r_0)$. In this sense, there are more threats to the qualification IV design than to the first-choice design.

Propositions 1 and 2 shed light on a contrast between the qualification IV and the first-choice research designs. Unlike the first-choice research design, strategy-proofness for schools is no longer sufficient for the qualification IV research design to extract a random assignment. It may extract an unintended broken random assignment not only for the DA or top trading cycles mechanism but also for the Boston mechanism and random serial dictatorship.

5 Discussion

5.1 Alternative Definition of a Random Assignment

My analysis of the first-choice design is based on Definition 2 of "random assignment." This definition requires that all students in realized fixed set $First_s(r_0)$ share the same assignment probability (propensity score). A possible alternative definition treats $First_s(R)$ as random and requires that

$$P(D_{is}(R) = 1 | i \in First_s(R), \theta_i = \theta) = P(D_{is}(R) = 1 | i \in First_s(R))$$

for all i for whom these conditional probabilities are defined. R denotes the random (not realized) lottery number profile. An equivalent property is

$$P(D_{js}(R) = 1 | j \in First_s(R)) = P(D_{ks}(R) = 1 | k \in First_s(R))$$

for all j and k for whom these conditional probabilities are defined. This alternative definition requires that treatment school assignment $D_{is}(R)$ is independent of type θ_i or the propensity score as a confounder conditional on random event $i \in First_s(R)$. This independence conditional on a random event or statistic is reminiscent of Chamberlain (1980) and Rosenbaum (1984)'s conditional logit panel frameworks, where the treatment distribution is independent of individual heterogeneity conditional on the random empirical frequency of being treated in the past.

All of my arguments extend to this alternative definition. See Appendix A.1 (especially Remark 1) for why Theorem 1 is correct even under the alternative definition. The discussion about the Boston mechanism under Example 1 in Section 3.3 goes through even under the alternative definition since $P(D_{iA}(R) = 1 | i \in First_A(R), \theta_i = \theta) = 1/2$ for $i = 1, 2$ and is independent of θ. The analysis of the DA and Charlotte mechanisms in Example 1 in Sections 3.1 and 3.4 also remains the same since in the example,

$$P(D_{iA}(R) = 1 | i \in First_A(R), \theta_i = \theta_1) = 1 \neq 0 = P(D_{iA}(R) = 1 | i \in First_A(R), \theta_i = \theta_2),$$

where θ_1 and θ_2 denote student types having $\succ_1$ and $\succ_2$, respectively. This shows that the first-choice design does not extract a random assignment even according to the alternative definition.

5.2 Alternative Definitions of Research Designs

Constant Cutoff Qualification IV

Section 4 shows a potential problem with the qualification IV defined as $Z_{is}(r) \equiv 1\{\rho^{\varphi}_{is} + r_{is} \leq \max\{\rho^{\varphi}_{js} + r_{js} | D_{js}(r) = 1\}\}$, where $\max\{\rho^{\varphi}_{js} + r_{js} | D_{js}(r) = 1\}$ is a random priority cutoff that varies as the lottery outcome changes. This problem may be expected to be solved by a modification of the qualification IV. For any constant $\pi \in \mathbb{R}$, define the **constant cutoff qualification IV** by

$$Z^{\pi}_{is}(r) \equiv 1\{\rho^{\varphi}_{is} + r_{is} \leq \pi\}.$$

In practice, the econometrician would define constant π as the realized priority cutoff at school s, that is, $\pi \equiv \max\{\rho^{\varphi}_{js} + r_{0js} | D_{js}(r_0) = 1\}$ where r_0 is the realized lottery numbers in the data. The constant cutoff qualification IV trivially extracts a random assignment since

$$\begin{aligned}
&P(Z^{\pi}_{is}(R) = 1 | \rho^{\varphi}_{is} = \rho, \theta_i = \theta) \\
&= P(\rho^{\varphi}_{is} + r_{is} \leq \pi | \rho^{\varphi}_{is} = \rho, \theta_i = \theta) \\
&= P(r_{is} \leq \pi - \rho^{\varphi}_{is} | \rho^{\varphi}_{is} = \rho, \theta_i = \theta) \\
&= \pi - \rho,
\end{aligned}$$

which is independent of θ conditional on $\rho^{\varphi}_{is} = \rho$.

However, the constant cutoff qualification IV entails new problems apart from randomness. First, when using the realized priority cutoff $\pi \equiv \max\{\rho^{\varphi}_{js} + r_{0js} | D_{js}(r_0) = 1\}$, I define or select an instrument depending on the realized data. Such data-dependent model selection often makes standard statistical inference invalid (Leamer, 1978). In addition, perhaps more importantly, the constant cutoff qualification IV may violate other requirements for a valid IV than independence. It is about the "monotonicity" requirement for an IV. Let me consider this issue with the following example.

Example 3. There are applicants 1, 2, 3, 4 and schools A and B with the following preferences and priorities:

$$
\begin{aligned}
&\succ_1, \succ_2 : A, B, \emptyset \\
&\succ_3, \succ_4 : B, \emptyset \\
&\rho_A, \rho_B : \{1, 2, 3, 4\}.
\end{aligned}
$$

The capacity of each school is 1. The treatment school is A. Without loss of generality, let $\rho^{\varphi}_{1A} = \rho^{\varphi}_{2A} = 0$.

Consider two lottery outcomes at school A:

- $r_A \equiv (r_{1A}, r_{2A}, r_{3A}, r_{4A})$ with $r_{1A} < r_{2A} < r_{3A}, r_{4A}$ and $r_{2A} \leq 0.5$
- $r'_A \equiv (r'_{1A}, r'_{2A}, r'_{3A}, r'_{4A})$ with $r'_{3A}, r'_{4A} < r'_{2A} < r'_{1A}$ and $r'_{2A} > 0.5$.

Fix any lottery numbers r_B at school B. Since $r_{1A} < r_{2A} \leq 0.5$ and $0.5 < r'_{2A} < r'_{1A}$, by definition $Z^{0.5}_{1A}(r_A, r_B) = Z^{0.5}_{2A}(r_A, r_B) = 1$ and $Z^{0.5}_{1A}(r'_A, r_B) = Z^{0.5}_{2A}(r'_A, r_B) = 0$. On the other hand, for any gDA mechanism, $D_{1A}(r_A, r_B) = 1, D_{1A}(r'_A, r_B) = 0, D_{2A}(r_A, r_B) = 0$, and $D_{2A}(r'_A, r_B) = 1$. This violates the monotonicity requirement for $Z^{0.5}_{iA}$ as an instrument for D_{iA}: Endogenous treatment variables D_{1A} and D_{2A} move in the opposite directions in response to the same change in the IV from $Z^{0.5}_{iA} = 1$ to $Z^{0.5}_{iA} = 0$. Monotonicity is required by many modern IV models with heterogeneous behavior and treatment effects (Heckman and Vytlacil (2007) chapter 4, Manski (2008) chapter 3, Angrist and Pischke (2009) section 4.4). Therefore, while the constant cutoff qualification IV always extracts a random assignment, it may not be able to identify a causal effect due to monotonicity violations.

Furthermore, since $First_A(r) = \{1, 2\}$ for all r in Example 3, this monotonicity violation persists even if I restrict the sample to $First_A(r_0)$. Also, since both $\rho^{\varphi}_{1A} = \rho^{\varphi}_{2A}$ and $\rho_{1A} = \rho_{2A}$, the counterexample goes through even if I use original priorities to define an alternative constant cutoff qualification IV as $Z^{0.5}_{is}(r) \equiv 1\{\rho_{iA} + r_{is} \leq 0.5\}$. Finally, note that 1 and

2 share the same priority at all schools. Thus the constant cutoff qualification IV research design may fail to satisfy monotonicity even if I modify it to the more restricted version that conditions on having the same priority at all schools.

Constant Rank Qualification IV

Example 3 also shows that yet another potential modification of the qualification IV does not extract a random assignment. For any positive integer m, define the **constant rank qualification IV** by

$$Z_{is}^{m\text{-th}}(r) \equiv 1\{r_{is} \leq m\text{-th}(\{r_{js}|j \in I\})\},$$

where m-th$(\cdot)$ is the m-th order statistic.[26] The constant rank qualification IV extracts a random assignment since $P(Z_{is}^{m\text{-}th}(R) = 1|\rho_{is}^{\varphi} = \rho, \theta_i = \theta) = m/|I|$, which is independent of θ. However,

- $Z_{1A}^{2nd}(r_A, r_B) = Z_{2A}^{2nd}(r_A, r_B) = 1(= Z_{1A}^{0.5}(r_A, r_B) = Z_{2A}^{0.5}(r_A, r_B))$ and
- $Z_{1A}^{2nd}(r'_A, r_B) = Z_{2A}^{2nd}(r'_A, r_B) = 0(= Z_{1A}^{0.5}(r'_A, r_B) = Z_{2A}^{0.5}(r'_A, r_B))$.

Therefore, potential IV Z_{iA}^{2nd} violates monotonicity by the same reason for $Z_{iA}^{0.5}$.[27] The above discussion also shows that the simplest possible IV, the random number r_i itself, suffers from the same monotonicity violation.

Conditioning on the Priorities at All Schools

Going back to the original first-choice and qualification IV research designs, they might fail to extract a random assignment even if I modify them to the more refined version that conditions on sharing the same priority at *every* school. Consider the following modification of Example 1.

Example 4. There are applicants 1, 2, 3, and schools A and B with the following preferences and priorities:

$\succ_1 : A, B, \emptyset$
$\succ_2 : A, \emptyset$

[26] Other possible definitions include $Z_{is}^{m\text{-th}}(r) \equiv 1\{\rho_{is}^{\varphi} + r_{is} \leq m\text{-}th(\{\rho_{js}^{\varphi} + r_{js}|j \in I, \rho_{js}^{\varphi} = \rho_{is}^{\varphi}\})\}$ and $1\{\rho_{is}+r_{is} \leq m\text{-th}(\{\rho_{js}+r_{js}|j \in I, \rho_{js} = \rho_{is}\})\}$. The discussion below applies to these alternative definitions too.

[27] All of the above points in this section apply to the top trading cycles mechanism, as shown in the next Section 5.3.

$\succ_3 : B, A, \emptyset$
$\rho_A : 3, \{1, 2\}$
$\rho_B : \{1, 2, 3\},$

where student 3's priority at school A is walk zone priority. The indifferences in the school priorities are broken by STB. The capacity of each school is 1. The treatment school is A.

The only difference from Example 1 is ρ_B: School B is now indifferent among all students. In Example 4, students 1 and 2 rank A first and share the same priority at both A and B. However, students 1 and 2 do not share the same assignment probability at A for the DA or Charlotte mechanism. Under the DA mechanism,

$$First_A(r) = \begin{cases} \emptyset & \text{if } r_2 < r_1 < r_3 \\ \{1, 2\} & \text{otherwise,} \end{cases}$$

where r_i is student i's lottery number used by both schools. Nevertheless, enumerating all lottery outcomes shows that for the treatment school assignment D_{is} and the qualifiation IV Z_{is},

$$\begin{aligned} &P(D_{iA}(R) = 1 | \theta_i = \theta_1) = P(Z_{iA}(R) = 1 | \theta_i = \theta_1) = 1/2 \\ &\neq 1/3 = P(D_{iA}(R) = 1 | \theta_i = \theta_2) = P(Z_{iA}(R) = 1 | \theta_i = \theta_2). \end{aligned}$$

Thus neither the first-choice or qualification IV research design extracts a random assignment for the DA mechanism even if they condition on sharing the same priority at every school. For the first-choice research design, this point remains true even if using the alternative random assignment criterion in Section 5.1 since $P(D_{iA}(R) = 1 | i \in First_A(R), \theta_i = \theta_1) = 3/5 \neq 2/5 = P(D_{iA}(R) = 1 | i \in First_A(R), \theta_i = \theta_2)$.

Conditioning on the Whole Preference List

Likewise, conditioning on the whole preference list is not a solution. Consider yet another modification of Example 1.

Example 5. There are applicants 1, 2, 3, and schools A and B with the following preferences and priorities:

$\succ_1 : A, B, \emptyset$
$\succ_2 : A, B, \emptyset$

$\succ_3: B, A, \emptyset$
$\rho_A: 3, \{1,2\}$
$\rho_B: 1, 3, 2,$

where student 3's priority at school A and 1's priority at B are walk zone priority. The capacity of each school is 1. The treatment school is A.

In this example, students 2 shares the same preference list with 1, i.e., $\succ_1 = \succ_2$. Nevertheless, the first-choice and qualification IV research designs do not extract a random assignment for the DA or Charlotte mechanism. Student 1 is assigned to A when 1 has a better lottery number than 2 at A. Otherwise, 3 is assigned to A. Each of the two cases occurs with equal probability 1/2. Thus,

$$P(D_{1A}(R) = 1) = P(Z_{1A}(R) = 1) = 1/2 \neq 0 = P(Z_{2A}(R) = 1) = P(D_{2A}(R) = 1),$$

despite having

$$First_A(r) = \begin{cases} \{1,2\} & \text{if } r_{1A} < r_{2A} \\ \emptyset & \text{otherwise.} \end{cases}$$

Therefore, the first-choice or qualification IV research design does not necessarily extract a random assignment for the DA or Charlotte mechanism even if one additionally conditions on the entire preference list and requires every student to rank all schools. This means that the potential problem with the first-choice design is not driven by short preferences.

5.3 Top Trading Cycles Mechanism

Some cities such as New Orleans and San Francisco have used a mechanism outside the generalized DA class studied in this paper. This mechanism, the top trading cycles mechanism, is also advocated by researchers as a Pareto efficient mechanism that is strategy-proof for *students* (Abdulkadiroğlu and Sönmez, 2003). This mechanism is defined as follows.

Definition 6. The **top trading cycles (TTC) mechanism** creates a matching through the following procedure. Take any assignment problem as given.

(1) Draw lottery number r according to a lottery regime (STB or MTB).

(2) Define school s's ex post strict priority order $\succ_{r_s}$ over students by $i \succ_{r_s} i'$ if $\rho_{is} + r_{is} < \rho_{i's} + r_{i's}$.

(3) Given $\succ_I$ and $(\succ_{r_s})_{s\in S}$, run the following **top trading cycles algorithm** (Shapley and Scarf, 1974).

- Step $t \geq 1$: Each student i points to her most preferred acceptable remaining school (if any). Students who do not point to any school are assigned to $\emptyset$. Each school s points to its most preferred student. As there are a finite number of schools and students, there exists at least one cycle, i.e., a sequence of distinct schools and students $(i_1, s_1, i_2, s_2, \ldots, i_L, s_L)$ such that student i_1 points to school s_1, school s_1 points to student i_2, student i_2 points to school s_2, ..., student i_L points to school s_L, and, finally, school s_L points to student i_1. Every student i_l $(l = 1, \ldots, L)$ in any cycle is assigned to the school she is pointing to. Any student who has been assigned a school seat or the outside option as well as any school s which has been assigned students such that the number of them is equal to its capacity c_s is removed. If no student remains, the algorithm terminates. Otherwise, it proceeds to the next step.

This algorithm terminates in a finite number of steps because at least one student is matched with a school (or the outside option) at each step and there are only a finite number of students.

It is possible to apply the first-choice research design to data from the TTC mechanism. However, the TTC implementation of the first-choice research design turns out not to extract a random assignment, as stated in Proposition 1. In Example 1, the TTC mechanism always assigns 1 to A under all lottery realizations r (regardless of whether the lottery structure is STB or MTB); no randomization occurs. Therefore, $First_A(r) = \{1, 2\}$ for all r, but $P(D_{is}(R) = 1|\theta_i = \theta_1) = 1 \neq 0 = P(D_{is}(R) = 1|\theta_i = \theta_2)$. Thus the first-choice research design does not extract a random assignment for the TTC mechanism.[28]

This failure of the first-choice research design is related to the fact that the TTC mechanism is not strategy-proof for schools. In Example 1, imagine A's true preference is $3 \succ_A 2 \succ_A 1$ while B's is $1 \succ_B 2 \succ_B 3$. Under these true preferences, A is matched with 1. If A misreports $2 \succ'_A 1, 3$, however, A is matched with 2, who is preferred to 1 under $\succ_A$. Therefore, the TTC mechanism is not strategy-proof for A in Example 1. This provides further support for the practical necessity of strategy-proofness for schools for successful randomization under the first-choice research design.

The same example also implies that the qualification IV research design may fail for the TTC mechanism. In fact, the qualification IV research design for the TTC mechanism does

[28] This point remains the same under the alternative definition of extracting a random assignment in Section 5.1: $P(D_{iA}(R) = 1|i \in First_A(R), \theta_i = \theta_1) = 1 \neq 0 = P(D_{iA}(R) = 1|i \in First_A(R), \theta_i = \theta_2)$.

not extract a random assignment even without priorities. It can be seen from Example 2, where there are no priorities and the top trading cycles mechanism is equivalent to the DA mechanism (Pathak and Sethuraman, 2011).

5.4 Large Market Considerations

My analysis in Section 3 shows that the first-choice design may not extracts a random assignment for gDA mechanisms that are not strategy-proof for schools. There remains the puzzle, however, that even under the DA and Charlotte mechanisms, the first-choice design sometimes receives empirical support for randomization. In particular, some empirical applications find that in the first-choice subsample $First_s(r_0)$, observable covariates of students with $D_{is}(r_0) = 1$ are similar to covariates of students with $D_{is}(r_0) = 0$. This covariate balance test is a standard check of a necessary condition for randomization.[29] There appears to be a tension between their empirical findings and my theoretical result.

A potential resolution is hinted by Theorem 1, the sufficiency of strategy-proofness for schools. Unlike small counterexamples like Example 1, empirical work is only done with data with at least hundreds of students. Though the DA and Charlotte mechanisms are not strategy-proof for schools in general, these mechanisms are often approximately so in certain large markets with many students and schools, as has been shown empirically and theoretically (Roth and Peranson, 1999; Immorlica and Mahdian, 2005; Kojima and Pathak, 2009; Azevedo and Budish, 2013; Lee, 2016; Ashlagi et al., 2016). The reason is that as the number of students and schools grows, chain reactions of rejections and applications at schools ranked below a manipulating school — which make the DA and Charlotte mechanisms manipulable in Example 1 — become less likely to come back to the manipulating school and benefit it. Existing empirical settings with hundred or thousands of students may therefore be subject to large market forces that make the DA and Charlotte mechanisms almost non-manipulable by schools. If so, Theorem 1 suggests the first-choice research design approximately extracts a random assignment even for the DA and Charlotte mechanisms.

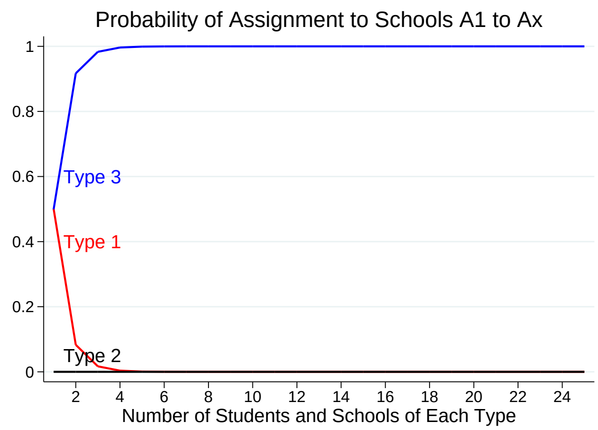
To see the effect of such large market forces, Figure 2 plots assignment probabilities for two types of expansions of Example 1 and the DA mechanism (equivalent to the Charlotte mechanism in this example).[30] A computer program to implement this simulation is available upon request. This figure reveals that as the market size grows, the discrepancy between student types 1 and 2's assignment probabilities at school A disappears, implying that breaks

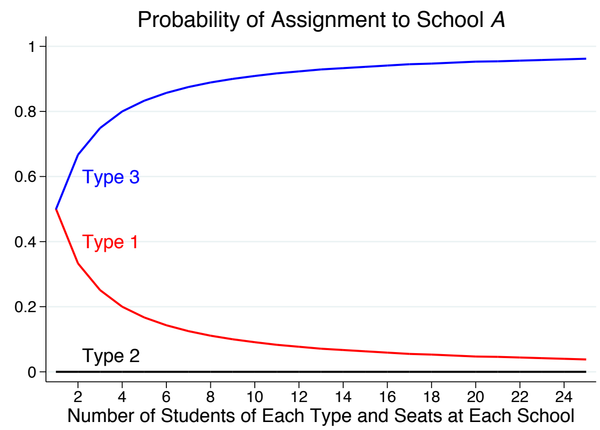
[29] See, for example, Hastings et al. (2009)'s Table VI, Deming (2011)'s Table I, Abdulkadiroğlu et al. (2014b)'s Table 2, Deming et al. (2014)'s Table A2.

[30] Figure 2b is of more interest in that its large market sequence is closer to the models of the above papers on strategy-proofness in large markets. For Figure 2a, Abdulkadiroğlu et al. (2017) complementarily show that in its limit, the first-choice design extracts a random assignment under the DA mechanism.

Figure 2: The Counterexample Evaporates in Large Markets

(a) Increasing Students and Number of Schools (b) Increasing Students and School *Size*





Notes: In Panel 2a, for each value of the x axis, I create an expansion of Example 1 with $2x$ schools $A_1, ..., A_x, B_1, ..., B_x$ with one seat each, and $3x$ students such that there are x students with each of the following three preferences:

$\succ_1: A_1, B_1, A_2, B_2, ..., A_x, B_x, \emptyset$
$\succ_2: A_1, ..., A_x, \emptyset$
$\succ_3: B_1, A_1, B_2, A_2, ..., B_x, A_x, \emptyset$
$\rho_{A_1}, ..., \rho_{A_x}: \{\text{students with } \succ_3\}, \{\text{students with } \succ_1 \text{ or } \succ_2\}$
$\rho_{B_1}, ..., \rho_{B_x}: \{\text{students with } \succ_1\}, \{\text{students with } \succ_2 \text{ or } \succ_3\}.$

In Panel 2b, for each value of the x axis, I create another expansion of Example 1 with x seats at each of schools A and B, and $3x$ students such that there are x students of each of the three types. For each scenario, I approximate the assignment probabilities by simulating the DA mechanism (equivalent to the Charlotte mechanism in this example) with STB 100000 times.

in randomization under the first-choice research design become smaller and smaller. This may explain why the first-choice research design often appears to extract a random assignment in empirical applications even for mechanisms that are not strategy-proof for schools. At the same time, the existing empirical support for the first-choice research design for the DA and Charlotte mechanisms (recall footnote 29) can be re-interpreted as suggesting that large market forces emphasized by the above theoretical papers are empirically relevant. Theorem 1 thus provides an asymptotic justification for the research design even for some mechanisms that are not strategy-proof for schools in general.[31]

[31] This suggests that many empirical studies using the first-choice design for non-strategy-proof mechanisms provide an unexpected set of empirical settings where treatment assignment is asymptotically random (but not exactly random in a finite sample). For inference, therefore, it is appropriate to use recent causal inference methods based on asymptotically random treatment assignment, such as Canay et al. (2014) and Belloni et al. (2014).

6 Conclusion

This paper provides a formal basis for understanding when and why the two popular empirical research designs do or do not extract a random assignment from school choice with lotteries. The first-choice design does so for mechanisms that are strategy-proof for schools. On the other hand, the qualification IV design does not necessarily extract a random assignment for any mechanism, thus suggesting a difference between the two research designs. Table 1 in the introduction provides a summary of the main results.

This paper takes a step toward deciphering theoretical structures hidden in empirical research designs exploiting market design with lotteries. This opens the door to several open questions. The most ambitious agenda is to design assignment mechanisms that enable as informative causal inference as possible (subject to welfare and strategic considerations). For example, it is intriguing to compare the Boston, DA, and top trading cycles mechanisms with different lottery regimes by their capabilities for quasi-experimental information production. The contrast between positive Corollary 1 and negative Proposition 1 is a step toward such a comparison. The empirically most important direction is to see if possible randomization failures (in Propositions 1 and 2) cause significant biases in treatment effect estimates in real data. I leave these challenging directions for future research.[32]

[32]There are many other directions of more technical nature. For example, while my framework assumes the use of random lottery numbers, some existing empirical studies use data with regression-discontinuity-style tie-breaking by admissions test scores. I would expect the point of the current paper to be valid even in regression discontinuity situations, but it is open how to extend this paper's results to a regression discontinuity setting. Also, my results point to the importance of strategy-proofness for schools within my mechanism class. It is thus important to characterize or axiomatize mechanisms that are strategy-proof for schools in the class. It is also a technical open question to use Theorem 1 and existing results on strategy-proofness in the large to formally justify the first-choice research design in large markets.

References

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

_ **, Josh Angrist, and Parag Pathak**, "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools," *Econometrica*, 2014, *82(1)*, 137–196.

_ **,** _ **, Yusuke Narita, and Parag A. Pathak**, "Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation," *Econometrica*, 2017.

_ **, Weiwei Hu, and Parag Pathak**, "Small High Schools and Student Achievement: Lottery-Based Evidence from New York City," 2014. Working Paper.

Agarwal, Nikhil and Paulo Somaini, "Demand Analysis using Strategic Reports: An Application to a School Choice Mechanism," 2015. Working Paper.

Ajayi, Kehinde, "Does School Quality Improve Student Performance? New Evidence from Ghana," 2013. Working Paper.

Angrist, Joshua D. and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, 2009.

Angrist, Joshua, Peter Hull, Parag Pathak, and Christopher Walters, "Leveraging Lotteries for School Value-Added: Testing and Estimation," *Quarterly Journal of Economics*, 2016. Forthcoming.

Ashlagi, Itai, Yash Kanoria, and Jacob Leshno, "Unbalanced Random Matching Markets: The Stark Effect of Competition," *Journal of Political Economy*, 2016, *125* (1), 69–98.

Azevedo, Eduardo and Eric Budish, "Strategyproofness in the Large," 2013. Working Paper.

Balinski, Michel and Tayfun Sönmez, "A Tale of Two Mechanisms: Student Placement," *Journal of Economic Theory*, 1999, *84*, 73–94.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, "High-Dimensional Methods and Inference on Structural and Treatment Effects," *Journal of Economic Perspective*, 2014, *28(2)*, 29–50.

Bergman, Peter, "The Effects of School Integration: Evidence from a Randomized Desegregation Program," 2016. Working Paper.

Beuermann, Diether, C Kirabo Jackson, and Ricardo Sierra, "Privately Managed Public Secondary Schools and Academic Achievement in Trinidad and Tobago: Evidence from Rule-Based Student Assignments," 2016. Working Paper.

Bloom, Howard S and Rebecca Unterman, "Can Small High Schools of Choice Improve Educational Prospects for Disadvantaged Students?," *Journal of Policy Analysis and Management*, 2014, *33* (2), 290–319.

Canay, Ivan A., Joseph P. Romano, and Azeem M. Shaikh, "Randomization Tests under an Approximate Symmetry Assumption," 2014. Working Paper.

Casella, George and Roger L Berger, *Statistical Inference*, Duxbury Pacific Grove, 2002.

Chamberlain, Gary, "Analysis of Covariance with Qualitative Data," *The Review of Economic Studies*, 1980, *47(1)*, 225–238.

Deming, David, "Better Schools, Less Crime?," *Quarterly Journal of Economics*, 2011, *126(4)*, 2063–2115.

_ **, Justine Hastings, Thomas Kane, and Douglas Staiger**, "School Choice, School Quality and Postsecondary Attainment," *American Economic Review*, 2014, *104(3)*, 991–1013.

Dobbie, Will and Roland G. Fryer, "Exam High Schools and Academic Achievement: Evidence from New York City," *American Economic Journal: Applied Economics*, 2014, *6(3)*, 58–75.

Ergin, Haluk and Tayfun Sönmez, "Games of School Choice under the Boston Mechanism," *Journal of Public Economics*, 2006, *90*, 215–237.

Fort, Margherita, Andrea Ichino, and Giulio Zanella, "Cognitive and Non-cognitive Costs of Daycare 0-2 for Girls," 2016. Working Paper.

Gale, David and Lloyd S. Shapley, "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, 1962, *69*, 9–15.

Hastings, Justine, Christopher Neilson, and Seth Zimmerman, "The Effect of School Choice on Intrinsic Motivation and Academic Outcomes," 2012. Working Paper.

_ , _ , and _ , “Are Some Degrees Worth More than Others? Evidence from College Admission Cutoffs in Chile,” 2013. Working Paper.

_ , Thomas J. Kane, and Douglas O. Staiger, “Heterogenous Preferences and the Efficacy of Public School Choice,” 2009. Working Paper.

Hatfield, John W., Fuhito Kojima, and Yusuke Narita, “Improving Schools Through School Choice: A Market Design Approach,” *Journal of Economic Theory*, 2016. (Forthcoming).

Heckman, James and Edward J. Vytlacil, “Econometric Evaluation of Social Programs Part II,” *Handbook of Econometrics*, 2007, *6(B)*, 4875–5143.

Immorlica, Nicole and Mohammad Mahdian, “Marriage, Honesty, and Stability,” *SODA*, 2005, pp. 53–62.

Jackson, Kirabo, “Do Students Benefit from Attending Better Schools? Evidence from Rule-based Student Assignments in Trinidad and Tobago,” *Economic Journal*, 2010, *120(549)*, 1399–1429.

_ , “Single-sex Schools, Student Achievement, and Course Selection: Evidence from Rule-based Student Assignments in Trinidad and Tobago,” *Journal of Public Economics*, 2012, *96(1-2)*, 173–187.

Kirkeboen, Lars, Edwin Leuven, and Magne Mogstad, “Field of Study, Earnings, and Self-Selection,” *Quarterly Journal of Economics*, 2016.

Kojima, Fuhito and Parag A. Pathak, “Incentives and Stability in Large Two-Sided Matching Markets,” *American Economic Review*, 2009, *99*, 608–627.

Leamer, Edward E., *Specification Searches: Ad Hoc Inference with Nonexperimental Data*, Wiley, 1978.

Lee, SangMok, “Incentive Compatibility of Large Centralized Matching Markets,” *Review of Economic Studies*, 2016. Forthcoming.

Lucas, Adrienne and Isaac Mbiti, “Effects of School Quality on Student Achievement: Discontinuity Evidence from Kenya,” *American Economic Journal: Applied Economics*, 2014, *6(3)*, 234–263.

Manski, Charles, *Identification for Prediction and Decision*, Cambridge: Harvard University Press, 2008.

McVitie, D. G. and L.B. Wilson, "Stable Marriage Assignments for Unequal Sets," *BIT*, 1970, *10*, 295–309.

Narita, Yusuke, "Match or Mismatch: Learning and Inertia in School Choice," 2015. Working Paper.

Pathak, Parag A and Jay Sethuraman, "Lotteries in Student Assignment: An Equivalence Result," *Theoretical Economics*, 2011, *6* (1), 1–17.

Pathak, Parag A. and Tayfun Sönmez, "Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism," *American Economic Review*, 2008, *98(4)*, 1636–1652.

Pop-Eleches, Cristian and Miguel Urquiola, "Going to a Better School: Effects and Behavioral Responses," *American Economic Review*, 2013, *103(4)*, 1289–1324.

Rosenbaum, Paul R., "Conditional Permutation Tests and the Propensity Score in Observational Studies," *Journal of the American Statistical Association*, 1984, pp. 565–574.

_ **and Donald B. Rubin**, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrica*, 1983, pp. 41–55.

Roth, Alvin E. and Elliott Peranson, "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design," *American Economic Review*, 1999, *89*, 748–780.

_ **and Marilda A. O. Sotomayor**, *Two-sided Matching: a Study in Game-theoretic Modeling and Analysis*, Cambridge University Press: Econometric Society monographs, 1992.

Roth, Alvin E and Marilda Sotomayor, "The College Admissions Problem Revisited," *Econometrica*, 1989.

Shapley, Lloyd and Herbert Scarf, "On Cores and Indivisibility," *Journal of Mathematical Economics*, 1974, *1*, 23–28.

Appendix

A Proofs

A.1 Proof of Theorem 1

Preliminaries

I start with lemmas to prove Theorem 1. I use $\varphi(r)$ as a shorthand for $\varphi(\rho, r)$, the assignment produced by a gDA mechanism φ when the reported priorities and lottery numbers are (ρ, r). As per usual, a **permutation** of r_s is a bijection from $\{r_{is}\}_{i\in I}$ to $\{r_{is}\}_{i\in I}$ itself.

Lemma 3. *Consider any assignment problem X, any lottery number profile r, and any gDA mechanism φ that is strategy-proof for schools. Let $\delta_s(r)$ be any permutation of r_s that switches only i' and i'' such that $\rho^{\varphi}_{i's} = \rho^{\varphi}_{i''s}$ and $\min\{\rho^{\varphi}_{i's} + r_{i's}, \rho^{\varphi}_{i''s} + r_{i''s}\} > \rho^{\varphi}_{is} + r_{is}$ for all i with $D_{is}(r) = 1$. That is, $\delta_{i's}(r) = r_{i''s}, \delta_{i''s}(r) = r_{i's}$, and $\delta_{js}(r) = r_{js}$ for all $j \neq i', i''$. If there are no such i' and i'', let $\delta_s(r)$ be the same as r_s. For any such $\delta_s(r)$, I have $\varphi(r) = \varphi(\delta_s(r), r_{-s})$.*

Proof of Lemma 3. For any assignment problem, a deterministic **assignment** (or a matching) is a vector μ that assigns each school s a set of at most c_s students $\mu_s \subset I$, and assigns each student i a seat at a school or the outside option $\mu_i \in S \cup \{\emptyset\}$. A matching μ is **individually rational** if $\mu_i \succeq_i \emptyset$ for every $i \in I$. With the notation $\rho^{\varphi} \equiv (\rho^{\varphi}_{is})_{i\in I, s\in S}$, μ is (ρ^{φ}, r)**-blocked** by $(s, i) \in S \times I$ if $s \succ_i \mu_i$ and either $|\mu_s| < c_s$ or there exists $\bar{i} \in I$ such that $i \succ^{\varphi}_{r_s} \bar{i}$, i.e., $\rho^{\varphi}_{is} + r_{is} < \rho^{\varphi}_{\bar{i}s} + r_{\bar{i}s}$. A matching μ is (ρ^{φ}, r)**-stable** if it is individually rational and not (ρ^{φ}, r)-blocked by (s, i). I use the following facts.

Fact 1. *(Roth and Sotomayor (1992)'s Theorem 5.8) For any assignment problem X, any lottery number profile r, any gDA mechanism φ, any (ρ^{φ}, r)-stable matching μ, any school s, any preference $\succ^{\varphi}_s$ responsive with respect to (c_s, ρ^{φ}, r), it holds $\mu_s \succeq^{\varphi}_s \varphi_s(\rho, r)$, where $\varphi_s(\rho, r)$ is the set of students assigned to s in the outcome of φ under X and r.*

Fact 2. *(Roth and Sotomayor (1989)'s Theorem 4) For any assignment problem X, any lottery number profile r, and any gDA mechanism φ, let μ and μ' be (ρ^{φ}, r)-stable matchings with $\mu_s \succ^{\varphi}_s \mu'_s$ for some preference $\succ^{\varphi}_s$ responsive with respect to $(c_s, \rho^{\varphi}_s, r_s)$. Then, for any $i \in \mu_s$ and $i' \in \mu'_s \setminus \mu_s$, it holds $i \succ^{\varphi}_{r_s} i'$, i.e., $\rho^{\varphi}_{is} + r_{is} < \rho^{\varphi}_{i's} + r_{i's}$.*

Fact 3. *Under any assignment problem, any lottery number profile r, any gDA mechanism φ, any $\delta_s(r)$ satisfying the conditions in the statement of Lemma 3, every (ρ^{φ}, r)-stable matching is also $(\rho^{\varphi}, (\delta_s(r), r_{-s}))$-stable.*

Proof of Fact 3. Let μ be a (ρ^φ, r)-stable matching. $i', i'' \notin \mu_s$ by Facts 1 and 2 and the assumption that $i', i'' \notin \varphi_s(\rho, r)$. The only potential $(\rho^\varphi, (\delta_s(r), r_{-s}))$-blocking pairs against μ are (s, i') and (s, i'') since the only difference between r and $(\delta_s(r), r_{-s})$ is the positions of i' and i'' in r_s. Since (1) both μ and $\varphi(\rho, r)$ are (ρ^φ, r)-stable and (2) $\min\{\rho^\varphi_{i's} + r_{i's}, \rho^\varphi_{i''s} + r_{i''s}\} > \rho^\varphi_{is} + r_{is}$ for all i with $D_{is}(r) = 1$, Facts 1 and 2 imply that for any $i \in \mu_s$, $\rho^\varphi_{is} + r_{is} < \min\{\rho^\varphi_{i's} + r_{i's}, \rho^\varphi_{i''s} + r_{i''s}\} = \min\{\rho^\varphi_{i's} + \delta_{i's}(r), \rho^\varphi_{i''s} + \delta_{i''s}(r)\}$, where the last equality comes from $\rho^\varphi_{i's} = \rho^\varphi_{i''s}$. Neither (s, i') nor (s, i'') thus $(\rho^\varphi, (\delta_s(r), r_{-s}))$-blocks μ. $\square$

Fact 4. *Under any assignment problem, any lottery number profile r, any gDA mechanism φ that is strategy-proof for schools, any $\delta_s(r)$ satisfying the conditions in the statement of Lemma 3, every $(\rho^\varphi, (\delta_s(r), r_{-s}))$-stable matching is also (ρ^φ, r)-stable.*

Proof of Fact 4. Suppose to the contrary that μ is $(\rho^\varphi, (\delta_s(r), r_{-s}))$-stable but not (ρ^φ, r)-stable. I show that it contradicts the assumption that φ is strategy-proof for schools.

Step 4.1. Either student i' or i'' (but not both) is in μ_s.

Proof of Step 4.1. Case i: Suppose that neither i' nor i'' (but not both) is in μ_s. By the $(\rho^\varphi, (\delta_s(r), r_{-s}))$-stability of μ, neither (s, i') or (s, i'') does $(\rho^\varphi, (\delta_s(r), r_{-s}))$-block μ. Either (s, i') or (s, i'') does (ρ^φ, r)-blocks μ, since (1) μ is $(\rho^\varphi, (\delta_s(r), r_{-s}))$-stable and so individually rational and (2) the only difference between r and $(\delta_s(r), r_{-s})$ is the positions of i' and i'' in r_s. That is, there exists $i \in \mu_s$ such that $\min\{\rho^\varphi_{i's} + r_{i's}, \rho^\varphi_{i''s} + r_{i''s}\} < \rho^\varphi_{is} + r_{is} < \min\{\rho^\varphi_{i's} + \delta_{i's}(r), \rho^\varphi_{i''s} + \delta_{i''s}(r)\} = \min\{\rho^\varphi_{i's} + r_{i's}, \rho^\varphi_{i''s} + r_{i''s}\}$, a contradiction. *Case ii*: Suppose that both i' and i'' are in μ_s. Since μ is not (ρ^φ, r)-stable but is individually rational (by its $(\rho^\varphi, (\delta_s(r), r_{-s}))$-stability), there exists i such that $s \succ_i \mu_i$ and $\rho^\varphi_{is} + r_{is} < \max\{\rho^\varphi_{i's} + r_{i's}, \rho^\varphi_{i''s} + r_{i''s}\} = \max\{\rho^\varphi_{i's} + \delta_{i's}(r), \rho^\varphi_{i''s} + \delta_{i''s}(r)\}$, a contradiction to the $(\rho^\varphi, (\delta_s(r), r_{-s}))$-stability of μ. $\square$

Step 4.2. $\varphi_s(r) \neq \varphi_s(\delta_s(r), r_{-s})$.

Proof of Step 4.2. $i', i'' \notin \varphi_s(r)$ (by assumption), either i' or i'' is in μ_s (by Step 3.1), and Fact 3 implies that (ρ^φ, r)-stable $\varphi_s(r)$ is $(\rho^\varphi, (\delta_s(r), r_{-s}))$-stable. Thus, Facts 1 and 2 implies that for any $\succ^{\varphi\delta}_s$ responsive with respect to $(c_s, \rho^\varphi_s, (\delta_s(r), r_{-s}))$, it holds that $\varphi_s(r) \succ^{\varphi\delta}_s \mu_s$, where I use the fact that μ is not (ρ^φ, r)-stable and so $\mu \neq \varphi(r)$. Fact 1 implies that for any $\succ^{\varphi\delta}_s$ responsive with respect to $(c_s, \rho^\varphi_s, (\delta_s(r), r_{-s}))$, it holds that $\mu_s \succeq^{\varphi\delta}_s \varphi_s(\delta_s(r), r_{-s})$. The two preference relations $\varphi_s(r) \succ^{\varphi\delta}_s \mu_s$ and $\mu_s \succeq^{\varphi\delta}_s \varphi_s(\delta_s(r), r_{-s})$ jointly imply that for any $\succ^{\varphi\delta}_s$ responsive with respect to $(c_s, \rho^\varphi_s, (\delta_s(r), r_{-s}))$, $\varphi_s(r) \succ^{\varphi\delta}_s \varphi_s(\delta_s(r), r_{-s})$, implying $\varphi_s(r) \neq \varphi_s(\delta_s(r), r_{-s})$. $\square$

Step 4.3. $\varphi_s(r) \succ_s^{\varphi} \varphi_s(\delta_s(r), r_{-s})$ for some preference $\succ_s^{\delta}$ responsive with respect to $(c_s, \rho_s, \delta_s(r))$, a contradiction to the assumption that φ is strategy-proof for schools.

Proof of Step 4.3. Take any $\succ_s$ responsive with respect to (c_s, ρ_s, r_s). If $\varphi_s(\delta_s(r), r_{-s}) \succ_s \varphi_s(r)$, then it is a contradiction to the assumption that φ is strategy-proof for schools. This implies $\varphi_s(r) \succ_s \varphi_s(\delta_s(r), r_{-s})$ since $\varphi_s(r) \neq \varphi_s(\delta_s(r), r_{-s})$ as shown in Step 3.2. Note that $\succ_s$ is also responsive with respect to $(c_s, \rho_s, \delta_s(r))$ since (1) the only difference between r and $(\delta_s(r), r_{-s})$ is the positions of i' and i'' in r_s, and (2) $i', i'' \notin \varphi_s(r)$ by assumption. Therefore, $\varphi_s(r) \succ_s^{\varphi} \varphi_s(\delta_s(r), r_{-s})$ for some $\succ_s^{\delta}$ responsive with respect to $(c_s, \rho_s, \delta_s(r))$. □

□

Facts 3 and 4 imply that under any assignment problem, any lottery number profile r, any gDA mechanism φ that is strategy-proof for schools, any $\delta_s(r)$ satisfying the conditions in the statement of Lemma 3, the set of (ρ^{φ}, r)-stable matchings coincides with the set of $(\rho^{\varphi}, (\delta_s(r), r_{-s}))$-stable matchings. Each student or school except s has the same set of responsive preferences over these common stable matchings both under r and $(\delta_s(r), r_{-s})$. School s also has the same preference over these stable matchings both under r and $(\delta_s(r), r_{-s})$ since (1) the only difference between r and $(\delta_s(r), r_{-s})$ is the positions of i' and i'' in r_s, and (2) $\min\{\rho^{\varphi}_{i's} + r_{i's}, \rho^{\varphi}_{i''s} + r_{i''s}\} > \rho^{\varphi}_{is} + r_{is}$ for all i with $D_{is}(r) = 1$ and thus $D_{i's}(r) = D_{i''s}(r) = 0$, which in turn implies $i', i'' \notin \mu_s$ for any (ρ^{φ}, r)- or $(\rho^{\varphi}, (\delta_s(r), r_{-s}))$-stable matching μ. Therefore, the school pessimal (ρ^{φ}, r)-stable matching is the same as the school pessimal $(\rho^{\varphi}, (\delta_s(r), r_{-s}))$-stable matching, i.e., $\varphi(r) = \varphi(\delta_s(r), r_{-s})$, proving Lemma 3. □

Lemma 4. *Consider any assignment problem X, any lottery number profile r, and any gDA mechanism φ. Let $\delta_s(r)$ be any permutation of r_s that switches only two students i' and i'' such that $\rho^{\varphi}_{i's} = \rho^{\varphi}_{i''s}$ and there exists i with $D_{is}(r) = 1$ such that $\max\{\rho^{\varphi}_{i's} + r_{i's}, \rho^{\varphi}_{i''s} + r_{i''s}\} \leq \rho^{\varphi}_{is} + r_{is}$. If there are no such i' and i'', let $\delta_s(r)$ be the same as r_s. For any such $\delta_s(r)$, I have $\varphi(r) = \varphi(\delta_s(r), r_{-s})$.*

Proof of Lemma 4. Under r or $(\delta_s(r), r_{-s})$, let t_0 be the step in the DA algorithm at which either i' or i'' or both first apply for s. If there is no such a step t_0 under either r or $(\delta_s(r), r_{-s})$, then the DA algorithm works in the same way until its end both under r and $(\delta_s(r), r_{-s})$, completing the proof. Assume the existence of such a step t_0 under both r and $(\delta_s(r), r_{-s})$. Until step $t_0 - 1$, the DA algorithm operates in the same way both under r and $(\delta_s(r), r_{-s})$ since the only difference between the two situations is the positions of i' and i'' in r_s. t_0 is thus common to r and $(\delta_s(r), r_{-s})$. Let I_{st_0} be the set of students who are kept by s from step $t_0 - 1$ or newly apply for s in step t_0. I_{st_0} is again the same between r and $(\delta_s(r), r_{-s})$. There are a few cases to consider.

Case I: Both i' and i'' apply for s at step t_0. Under r, s tentatively accepts both i' and i'' by the assumption that there exists i with $D_{is}(r) = 1$ such that $\max\{\rho^{\varphi}_{i's} + r_{i's}, \rho^{\varphi}_{i''s} + r_{i''s}\} \leq \rho^{\varphi}_{is} + r_{is}$. Under $(\delta_s(r), r_{-s})$, s again tentatively accepts both i' and i''. This is because $\{\rho^{\varphi}_{is} + r_{is}\}_{i \in I_{st_0}} = \{\rho^{\varphi}_{is} + \delta_i(r_s)\}_{i \in I_{st_0}}$ (recall $i', i'' \in I_{st_0}$ by assumption and I_{st_0} is the same between r and $(\delta_s(r), r_{-s})$) and the above fact that s tentatively accepts both i' and i'' under r, which jointly imply that $\max\{\rho^{\varphi}_{i's} + \delta_{i's}(r), \rho^{\varphi}_{i''s} + \delta_{i''s}(r)\} = \max\{\rho^{\varphi}_{i's} + r_{i's}, \rho^{\varphi}_{i''s} + r_{i''s}\} \leq c_s$-th$(\{\rho^{\varphi}_{is} + r_{is}\}_{i \in I_{st_0}})$ where c_s-th$(\cdot)$ is the c_s-th order statistic. The DA algorithm also works in the same way for the remaining steps.

Case II: Only one of i' or i'' applies for s at step t_0. Without loss of generality, suppose only i' applies for s at step t_0. Under r, s tentatively accepts i' by the assumption that there exists i with $D_{is}(r) = 1$ such that $\rho^{\varphi}_{i's} + r_{i's} \leq \max\{\rho^{\varphi}_{i's} + r_{i's}, \rho^{\varphi}_{i''s} + r_{i''s}\} \leq \rho^{\varphi}_{is} + r_{is}$.

Case II.A: $r_{i's} > r_{i''s}$ (and so $\delta_{i's}(r) < \delta_{i''s}(r)$). Under $(\delta_s(r), r_{-s})$, s also tentatively accepts i' by the following reason. By the above fact that s tentatively accepts i' under r, it holds that $\rho^{\varphi}_{i's} + r_{i's} \leq c_s$-th$(\{\rho^{\varphi}_{is} + r_{is}\}_{i \in I_{st_0}})$, implying $\rho^{\varphi}_{i's} + \delta_{i's}(r) < \rho^{\varphi}_{i's} + r_{i's} \leq c_s$-th$(\{\rho^{\varphi}_{is} + \delta_{is}(r)\}_{i \in I_{st_0}})$.

Case II.B: $r_{i's} < r_{i''s}$. Under $(\delta_s(r), r_{-s})$, s also tentatively accepts i' by the following reason. Suppose not. Then c_s-th$(\{\rho^{\varphi}_{is} + r_{is}\}_{i \in I_{st_0}}) \leq c_s$-th$(\{\rho^{\varphi}_{is} + \delta_i(r_s)\}_{i \in I_{st_0}}) < \rho^{\varphi}_{i's} + \delta_{i's}(r) = \rho^{\varphi}_{i''s} + r_{i''s}$, where the first inequality $r_{i's} < r_{i''s}$ and the last equality uses the assumption $\rho^{\varphi}_{i's} = \rho^{\varphi}_{i''s}$. Let's call c_s-th$(\{\rho^{\varphi}_{is} + r_{is}\}_{i \in I_{st_0}})$ the *tentative cutoff for school s at step* t_0, which is common between r and $(\delta_s(r), r_{-s})$ since I_{st_0} is the same between r and $(\delta_s(r), r_{-s})$ and $r_{is} = \delta_{is}(r)$ for all $i \in I_{st} \setminus i'$. Since the tentative cutoff is monotonically decreasing in steps, the above inequality implies that for all i with $D_{is}(r) = 1$, $\rho^{\varphi}_{is} + r_{is} < \rho^{\varphi}_{i's} + \delta_{i's}(r) = \rho^{\varphi}_{i''s} + r_{i''s}$, contradicting the assumption that there exists i with $D_{is}(r) = 1$ such that $\rho^{\varphi}_{i''s} + r_{i''s} \leq \max\{\rho^{\varphi}_{i's} + r_{i's}, \rho^{\varphi}_{i''s} + r_{i''s}\} \leq \rho^{\varphi}_{is} + r_{is}$.

In all cases, the DA algorithm works in the same way at step t_0 under r and $(\delta_s(r), r_{-s})$. In Case I, the DA algorithm also works in the same way for the remaining steps. In Case II, let t_1 be the step in the DA algorithm at which i'' first applies for s. If there is no such a step t_1 under either r or $(\delta_s(r), r_{-s})$, then the DA algorithm works in the same way until its end both under r and $(\delta_s(r), r_{-s})$, completing the proof. Assume the existence of such a step t_1 under both r and $(\delta_s(r), r_{-s})$. Until step $t_1 - 1$, the DA algorithm operates in the same way both under r and $(\delta_s(r), r_{-s})$ since the only difference between the two situations is the positions of i' and i'' in r_s. t_1 is thus common to r and $(\delta_s(r), r_{-s})$. Let I_{st_1} be the set of students who are kept by s from step $t_1 - 1$ or newly apply for s in step t_1. I_{st_1} is again the same between r and $(\delta_s(r), r_{-s})$. I again consider the following two cases.

Case II.A (Continued): $r_{i's} > r_{i''s}$. Under $(\delta_s(r), r_{-s})$, s also tentatively accepts i'' by the following reason. Suppose not. Then $c_s\text{-th}(\{\rho^{\varphi}_{is} + r_{is}\}_{i\in I_{st_1}}) \leq c_s\text{-th}(\{\rho^{\varphi}_{is} + \delta_i(r_s)\}_{i\in I_{st_1}}) < \rho^{\varphi}_{i''s} + \delta_{i''s}(r) = \rho^{\varphi}_{i's} + r_{i's}$. Since the tentative cutoff is monotonically decreasing in steps, the above inequality implies that for all i with $D_{is}(r) = 1$, $\rho^{\varphi}_{is} + r_{is} < \rho^{\varphi}_{i's} + \delta_{i's}(r) = \rho^{\varphi}_{i''s} + r_{i''s}$, contradicting the assumption that there exists i with $D_{is}(r) = 1$ such that $\rho^{\varphi}_{i''s} + r_{i''s} \leq \max\{\rho^{\varphi}_{i's} + r_{i's}, \rho^{\varphi}_{i''s} + r_{i''s}\} \leq \rho^{\varphi}_{is} + r_{is}$.

Case II.B (Continued): $r_{i's} < r_{i''s}$. By the above fact that s tentatively accepts i'' under r, it holds that $\rho^{\varphi}_{i''s} + r_{i''s} \leq c_s\text{-th}(\{\rho^{\varphi}_{is} + r_{is}\}_{i\in I_{st_1}})$, implying $\rho^{\varphi}_{i''s} + \delta_{i''s}(r) < c_s\text{-th}(\{\rho^{\varphi}_{is} + \delta_{is}(r)\}_{i\in I_{st_1}})$.

In both cases, the DA algorithm works in the same way at step t_1 under r and $(\delta_s(r), r_{-s})$. Since both i_0 and i_1 have already applies for s by step t_1 or never apply for s, the DA algorithm also works in the same way for the remaining steps, showing Lemma 4. □

Main Proof

Suppose that the first-choice research design does not extract a random assignment for gDA mechanism φ for some school s at some assignment problem X. Fix φ, s, and X throughout. I show that this supposition implies that φ is not strategy-proof for schools. For each lottery number profile r, define $i_0(r)$ and $i_1(r)$ by two students who satisfy the following conditions.

- $i_0(r), i_1(r) \in First_s(r)$
- $D_{i_0(r)s}(r) = 0$ and $D_{i_1(r)s}(r) = 1$
- $r_{i_0(r)s} \leq r_{is}$ for all $i \in First_s(r)$ with $D_{is}(r) = 0$
- $r_{i_1(r)s} \geq r_{is}$ for all $i \in First_s(r)$ with $D_{is}(r) = 1$.

If there are no two students satisfying the conditions, let $i_0(r) = i_1(r) = \emptyset$. $i_0(r)$ and $i_1(r)$ are uniquely well-defined for all r. (If there are two $\hat{i}_0(r) \neq \tilde{i}_0(r)$ satisfying the conditions, then $r_{\hat{i}_0(r)s} > r_{\tilde{i}_0(r)s}$ and $r_{\hat{i}_0(r)s} < r_{\tilde{i}_0(r)s}$, a contradiction. The same logic holds for $i_1(r)$ too.) This proof uses the following equivalent representation of gDA mechanism φ.

Definition 7. Algorithm $\texttt{2STAGES}(r)$ operates in the following way.

(1) Same as in Definition 1.

(2) Same as in Definition 1.

(3) Given $\succ_I$ and $(\succ^{\varphi}_{r_s})_{s\in S}$, run the following sub-algorithm $\texttt{STAGE1}(r)$: Remove $i_0(r)$ and $i_1(r)$ from X (without changing anything else) and run the DA algorithm on the remaining subproblem where schools' strict priorities are given by $(\succ^{\varphi}_{r_s})$.

(4) Starting from the output of $\texttt{STAGE1}(r)$ as the initial tentative assignment, run the following sub-algorithm $\texttt{STAGE2}(r)$: Include $i_0(r)$ and $i_1(r)$ and run the DA algorithm where only $i_0(r)$ and $i_1(r)$ apply for any school and schools' strict priorities are given by $(\succ^{\varphi}_{r_s})$.

By McVitie and Wilson (1970)'s order irrelevance result, $\texttt{2STAGES}(r)$ and $\varphi(r)$ (the simplified notation for $\varphi(\rho, r)$) produce the same matching for all r. Let $t-1$ be the last step of $\texttt{STAGE1}(r)$ at which $\texttt{STAGE1}(r)$ stops and $\mu_{t-1}(r) \equiv (\mu_{st-1}(r))_{s\in S}$ be the tentative matching at the end of step $t-1$. Start counting $\texttt{STAGE2}(r)$'s steps without resetting the step index so that t is the initial step of $\texttt{STAGE2}(r)$. (Note that t implicitly depends on r.)

For each lottery number profile r, define $\sigma^*(r) = (\sigma^*_{s'}(r))_{s'\in S}$ as the following permutation of r. If $i_0(r) = i_1(r) = \emptyset$, then $\sigma^*_{is}(r) = r_{is}$ for all student i and school s. Otherwise, if gDA mechanism φ uses MTB, $\sigma^*(r)$ is obtained by switching only $i_0(r)$ and $i_1(r)$ only in r_s, i.e.,

- $\sigma^*_{i_0(r)s}(r) = r_{i_1(r)s}$ and $\sigma^*_{i_1(r)s}(r) = r_{i_0(r)s}$
- $\sigma^*_{is}(r) = r_{is}$ for all $i \neq i_0(r), i_1(r)$
- $\sigma^*_{is'}(r) = r_{is'}$ for all i and $s' \neq s$.

If φ uses STB, $\sigma^*(r)$ is obtained by switching $i_0(r)$ and $i_1(r)$ in $r_{s'}$ for all s', i.e., for all $s' \in S$, including s,

- $\sigma^*_{i_0(r)s'}(r) = r_{i_1(r)s'}$ and $\sigma^*_{i_1(r)s'}(r) = r_{i_0(r)s'}$
- $\sigma^*_{is'}(r) = r_{is'}$ for all $i \neq i_0(r), i_1(r)$.

Given any r, I say two students i' and i'' are **consecutive in** r_s **within** $First_s(r)$ if $i', i'' \in First_s(r)$ and there is no $i''' \in First_s(r)$ such that $r_{i's} < r_{i'''s} < r_{i''s}$ or $r_{i's} > r_{i'''s} > r_{i''s}$. For each r, consider any permutation $\hat{\sigma}_s(r) \neq \sigma^*_s(r)$ of r_s that switches only two students i' and i'' who are consecutive in r_s within $First_s(r)$, i.e., $\hat{\sigma}_{i's}(r) = r_{i''s}, \hat{\sigma}_{i''s}(r) = r_{i's}$, and $\hat{\sigma}_{js}(r) = r_{js}$ for all $j \neq i', i''$. If there are no such i' and i'', let $\hat{\sigma}_s(r) = r_s$. Let $\hat{\sigma}(r)$ be the following. If φ uses MTB, let $\hat{\sigma}(r) = (\hat{\sigma}_s(r), r_{-s})$. If φ uses STB, let $\hat{\sigma}(r) = \times_{|S|}\hat{\sigma}_s(r)$.

Lemma 5. *(A breakdown of the Fisher property discussed in Section 3.3) There exits lottery number profile r consistent with gDA mechanism φ's lottery structure (STB or MTB) satisfying the following: $D_{i_0(r)s}(\sigma^*(r)) = D_{i_1(r)s}(\sigma^*(r)) = 0$ or there exist some student $i \in First_s(r)$ and some permutation $\hat{\sigma}_s(r)$ defined above such that $D_{is}(\hat{\sigma}(r)) \neq D_{is}(r)$.*

Proof of Lemma 5. It is enough to show that if there is no r such that there exists $i \in First_s(r)$ such that $D_{is}(\hat{\sigma}(r)) \neq D_{is}(r)$, then there exits r such that $D_{i_0(r)s}(\sigma^*(r)) = D_{i_1(r)s}(\sigma^*(r)) = 0$. Suppose to the contrary that for all r, it is not the case $D_{i_0(r)s}(\sigma^*(r)) = D_{i_1(r)s}(\sigma^*(r)) = 0$.

Step 5.A. For all lottery number profile r with $i_0(r) \neq \emptyset$ and $i_1(r) \neq \emptyset$, $D_{i_0(r)s}(\sigma^*(r)) = 1, D_{i_1(r)s}(\sigma^*(r)) = 0$, and $D_{is}(\sigma^*(r)) = D_{is}(r)$ for all student $i \in First_s(r)$ with $i \neq i_0(r)$ and $i \neq i_1(r)$.

Proof of Step 5.A. By the above assumption that for all r, it is not the case $D_{i_0(r)s}(\sigma^*(r)) = D_{i_1(r)s}(\sigma^*(r)) = 0$, it is enough to show that for all such r, $D_{i_1(r)s}(\sigma^*(r)) = 0$ and $D_{is}(\sigma^*(r)) = D_{is}(r)$ for all $i \in First_s(r)$ with $i \neq i_0(r)$ and $i \neq i_1(r)$. Let me consider $\texttt{STAGE1}(r)$ and $\texttt{STAGE1}(\sigma^*(r))$. Since everything except $i_0(r)$ and $i_1(r)$'s lottery numbers is the same between r and $\sigma^*(r)$, both $\texttt{STAGE1}(r)$ and $\texttt{STAGE1}(\sigma^*(r))$ produce the same tentative assignment $\mu_{st-1}(r) = \mu_{st-1}(\sigma^*(r)) \equiv \mu_{st-1}$. Now start $\texttt{STAGE2}(r)$ and $\texttt{STAGE2}(\sigma^*(r))$. Under r, McVitie and Wilson (1970)'s order irrelevance result implies s rejects $i_0(r)$ and tentatively accepts $i_1(r)$, which implies $\rho^{\varphi}_{i_0(r)s} + r_{i_0(r)s} > c_s\text{-th}(\{\rho^{\varphi}_{is} + r_{is}\}_{i \in \mu_{st-1} \cup i_0(r) \cup i_1(r)})$ where $c_s\text{-th}(\cdot)$ is the c_s-th order statistic in the input set. Under $\sigma^*(r)$, by definition of $\sigma^*(r)$, $\rho^{\varphi}_{i_1(r)s} + \sigma^*_{i_1(r)s}(r) = \rho^{\varphi}_{i_0(r)s} + r_{i_0(r)s} > c_s\text{-th}(\{\rho^{\varphi}_{is} + r_{is}\}_{i \in \mu_{st-1} \cup i_0(r) \cup i_1(r)})$, resulting in s's rejecting $i_1(r)$. Since any rejected student is never be accepted in the DA algorithm, this implies $D_{i_1(r)s}(\sigma^*(r)) = 0$. $D_{is}(\sigma^*(r)) = D_{is}(r)$ for all $i \in First_s(r)$ with $i \neq i_0(r)$ and $i \neq i_1(r)$ by the following reason. Suppose not. There exists $i \in \mu_{st-1} \cap First_s(r) \setminus \{i_0(r), i_1(r)\}$ for whom, without loss of generality, $D_{is}(\sigma^*(r)) = 0$ and $D_{is}(r) = 1$. This implies that $D_{i_0(r)s}(\sigma^*(r)) = D_{i_1(r)s}(\sigma^*(r)) = 0$ since $\rho^{\varphi}_{is} + r_{is} < \rho^{\varphi}_{i_1(r)s} + r_{i_1(r)s} < \rho^{\varphi}_{i_0(r)s} + r_{i_0(r)s}$ (the first inequality is by definition of $i_1(r)$) and so $\rho^{\varphi}_{is} + \sigma^*_{is}(r) < \rho^{\varphi}_{i_0(r)s} + \sigma^*_{i_0(r)s}(r) < \rho^{\varphi}_{i_1(r)s} + \sigma^*_{i_1(r)s}(r)$. This is a contradiction to the assumption that for all r, it is not the case $D_{i_0(r)s}(\sigma^*(r)) = D_{i_1(r)s}(\sigma^*(r)) = 0$. □

Step 5.B. For all lottery number profile r with $i_0(r) = i_1(r) = \emptyset$, it holds $\varphi(\sigma^*(r)) = \varphi(r)$. For each r and each permutation $\hat{\sigma}(r) \neq \sigma^*(r)$ defined right before Lemma 5, $D_{is}(\hat{\sigma}(r)) = D_{is}(r)$ for all $i \in First_s(r)$.

Proof of Step 5.B. For all r with $i_0(r) = i_1(r) = \emptyset$, $\sigma^*(r) = r$ and it is trivial that $\varphi(\sigma^*(r)) = \varphi(r)$. The second part is by assumption. $\square$

Step 5.C. For each lottery number profile r, define $o_s(r) \equiv |\{i \in First_s(r)|D_{is}(r) = 1\}|$. For each r and each permutation $\sigma_s(r)$ of r_s that permutes lottery numbers only among members of $First_s(r)$, let $\sigma(r)$ be the following. If φ uses MTB, $\sigma(r) = (\sigma_s(r), r_{-s})$. If φ uses STB, $\sigma(r) = \times_{|S|}\sigma_s(r)$. Then the following is true for all r.

(1) *$First_s(\sigma(r)) = First_s(r)$.*

(2) *$o_s(\sigma(r)) = o_s(r)$.*

(3) *For all i and i' in $First_s(\sigma(r))$, if $D_{is}(\sigma(r)) > D_{i's}(\sigma(r))$, then $\sigma_{is}(r) < \sigma_{i's}(r)$.*

Proof of Step 5.C. Since any permutation can be expressed as a composition of contrapositions (permutations switching consecutive two elements), I can express any σ as a composition of σ^* and $\hat{\sigma}$'s defined right before Lemma 5. Steps 5.A and 5.B imply (1) and (2). (3) follows from the fact that for all r and all $i, i' \in First_s(r)$, $\rho^{\varphi}_{is} = \rho^{\varphi}_{i's}$ and another well-known property of the DA algorithm that for applicants who rank s first and share ρ^{φ}_{is}, D_{is} is monotonically decreasing in r_{is} (Balinski and Sönmez, 1999). $\square$

Let $\mathcal{R} \equiv \{r \in [0,1]^{|I|\times|S|}|r_{is} \neq r_{js}$ for all students i, j, and school $s\}$ be the set of all possible values of the lottery number profile r. Partition $\mathcal{R}$ into $\mathcal{P} \equiv \{\mathcal{R}_n\}_{n\in\mathcal{N}}$ ($\mathcal{N}$ is an uncountable set of indices) such that the following holds: Within each $\mathcal{R}_n$, for all $r, r' \in \mathcal{R}_n$, r can be obtained from r' by permutation $r' = \sigma(r)$, where $\sigma(r)$ is a permutation of r defined in Step 5.C. This partition is well-defined by Step 5.C(1): Since Step 5.C(1) guarantees $First_s(r) = First_s(r')$, $r' = \sigma(r)$ for such a permutation σ if and only if $r = \sigma(r')$ for such a (possibly different) permutation σ. Let r_n be a generic element of $\mathcal{R}_n$. Note that $First_s(r_n)$ and $o_s(r_n)$ are the same for all $r_n \in \mathcal{R}_n$ by Step 5.C(1) and 5.C(2), respectively. Step 5.C guarantees that conditional on each $\mathcal{R}_n$, $D_{is}(R)$ is independent of i's type for all $i \in First_s(r_n)$, i.e., for all n and θ, assuming the rule of $0/0 = 0$,

$$P(D_{is}(R) = 1|i \in First_s(r_n), R \in \mathcal{R}_n, \theta_i = \theta) = \frac{o_s(r_n)}{|First_s(r_n)|}1\{First_s(r_n) \neq \emptyset\},$$

which is independent of $\theta_i = \theta$. The equality holds since $First_s(r_n)$ and $o_s(r_n)$ stay constant across all $r_n \in \mathcal{R}_n$ and under each r_n, students with the $o_s(r_n)$-best lottery numbers have $D_{is}(r_n) = 1$ (by Step 5.C). Therefore, for each $n \in \mathcal{N}$,

$$
\begin{aligned}
&P(D_{is}(R)=1|i\in First_s(r_0),R\in\mathcal{R}_n,\theta_i=\theta)\\
&=\begin{cases}\dfrac{o_s(r_n)}{|First_s(r_n)|} & \text{if } \rho^{\varphi}_{is}=\rho^{\varphi}_{js} \text{ for any } j\in First_s(r_n)\neq\emptyset\\ 1 & \text{if } \rho^{\varphi}_{is}<\rho^{\varphi}_{js} \text{ for any } j\in First_s(r_n)\neq\emptyset\\ & \quad\text{or } (First_s(r_n)=\emptyset \text{ and } D_{ks}(r_n)=1 \text{ for all } k\in First_s(r_0))\\ 0 & \text{otherwise}\end{cases}\\
&\equiv p_n,
\end{aligned}
$$

where p_n is independent of θ_i conditional on $i \in First_s(r_0)$ since p_n depends on θ_i only through ρ^{φ}_{is}, but ρ^{φ}_{is} is pinned down by the fact that $i \in First_s(r_0)$.

Let Y be $Y(R)=\mathcal{R}^R_n$ where $\mathcal{R}^R_n$ is the element of the partition $\mathcal{P}$ with $R\in\mathcal{R}^R_n$. Let P_Y be the probability measure of Y induced by that of R, i.e. for all $A\subset\mathcal{N}, P_Y(\{\mathcal{R}_n\}_{n\in A})\equiv P(R\in\cup_{n\in A}\mathcal{R}_n)$. With this notation, assuming the rule of $0/0=0$,

$$
\begin{aligned}
&P(D_{is}(R)=1|i\in First_s(r_0),\theta_i=\theta)\\
&=\int_{\{\mathcal{R}_n\}_{n\in\mathcal{N}}} P(D_{is}(R)=1|i\in First_s(r_0),Y=\mathcal{R}_n,\theta_i=\theta)dP_Y(\mathcal{R}_n)\\
&\quad\text{(by the law of iterated expectation)}\\
&=\int_{\{\mathcal{R}_n\}_{n\in\mathcal{N}}} p_n dP_Y(\mathcal{R}_n),
\end{aligned}
$$

which is again independent of θ_i since both p_n and P_Y are independent of θ_i. Thus the first-choice research design extracts a random assignment, a contradiction. This completes the proof of Lemma 5. □

Remark 1. Under the alternative definition of a random assignment in Section 5.1, the last part of the proof of Lemma 5 simplifies to the following.

$$
\begin{aligned}
&P(D_{is}(R)=1|i\in First_s(R),\theta_i=\theta)\\
&=\int_{\{\mathcal{R}_n\}_{n\in\mathcal{N}}} P(D_{is}(R)=1|i\in First_s(R),Y=\mathcal{R}_n,\theta_i=\theta)dP_Y(\mathcal{R}_n)\\
&=\int_{\{\mathcal{R}_n\}_{n\in\mathcal{N}}} \frac{o_s(r_n)}{|First_s(r_n)|}1\{First_s(r_n)\neq\emptyset\}dP_Y(\mathcal{R}_n),
\end{aligned}
$$

which is independent of $\theta_i=\theta$. No other part of the proof depends on which random assignment definition I use. Therefore, Theorem 1 goes through even for the random assignment definition in Section 5.1.

Lemma 5 shows two possible scenarios. I consider these two cases, Cases 1 and 2 below, one by one.

Case 1: There exists lottery number profile r consistent with gDA mechanism φ's lottery structure (STB or MTB) such that $D_{i_0(r)s}(\sigma^*(r)) = D_{i_1(r)s}(\sigma^*(r)) = 0$. For each r, define r^*_s as the following permutation of r_s. If $i_0(r) = i_1(r) = \emptyset$, let $r^*_{is} = r_{is}$ for all student i and school s. Otherwise, let r^*_s be defined by the following conditions.

- $r^*_{is} = r_{is}$ for all i with $\rho^\varphi_{is} \neq \rho^\varphi_{i_0(r)s} = \rho^\varphi_{i_1(r)s}$
- $r^*_{i_1(r)s} = \max\{r_{is} | \rho^\varphi_{is} = \rho^\varphi_{i_0(r)s} = \rho^\varphi_{i_1(r)s}$ and $D_{is}(r) = 1\}$
- $r^*_{i_0(r)s} > r^*_{i_1(r)s}$ and there is no such i that $\rho^\varphi_{is} = \rho^\varphi_{i_0(r)s} = \rho^\varphi_{i_1(r)s}$ and $r^*_{i_0(r)s} > r^*_{is} > r^*_{i_1(r)s}$
- $r^*_{is} > r^*_{js}$ if and only if $r_{is} > r_{js}$ for all i, j such that $\rho^\varphi_{is} = \rho^\varphi_{js} = \rho^\varphi_{i_0(r)s} = \rho^\varphi_{i_1(r)s}$ and $i \neq i_0(r)$, $i \neq i_1(r)$, $j \neq i_0(r)$, and $j \neq i_1(r)$.

For each lottery number profile r and each school $s' \neq s$, define $\tilde{\sigma}_{s'}(r)$ as the following permutation of $r_{s'}$. If $i_0(r) = i_1(r) = \emptyset$ or MTB is used by φ, then $\tilde{\sigma}_{is'}(r) = r_{is'}$ for all student i. Otherwise, $\tilde{\sigma}_{s'}(r)$ is obtained by moving $i_1(r)$ to right above $i_0(r)$, i.e.,

- $\tilde{\sigma}_{i_1(r)s'}(r) = \max\{r_{is'} | r_{is'} < r_{i_0(r)s'}\}$
- $\tilde{\sigma}_{is'}(r) > \tilde{\sigma}_{js'}(r)$ if and only if $r_{is'} > r_{js'}$ for all students i, j such that $i, j \neq i_1(r)$.

Note that $\tilde{\sigma}_{s'}(r)$ implicitly depends on whole r.

Lemma 6. *(Outcome-equivalence between lottery number profiles r and $(r^*_s, \tilde{\sigma}_{-s}(r))$) For all lottery number profile r, I have $\varphi(r) = \varphi(r^*_s, \tilde{\sigma}_{-s}(r))$ where $\tilde{\sigma}_{-s}(r) \equiv (\tilde{\sigma}_{s'}(r))_{s' \neq s}$. Therefore, $D_{i_0(r)s}(r^*_s, \tilde{\sigma}_{-s}(r)) = 0 (= D_{i_0(r)s}(r))$ and $D_{i_1(r)s}(r^*_s, \tilde{\sigma}_{-s}(r)) = 1 (= D_{i_1(r)s}(r))$ whenever $i_0(r), i_1(r) \neq \emptyset$.*

Proof of Lemma 6. The following Steps 6.A and 6.B imply Lemma 6. □

Step 6.A. For all lottery number profile r, $\varphi(r) = \varphi(r^*_s, r_{-s})$.

Proof of Step 6.A. If $i_0(r) = i_1(r) = \emptyset$ and so $r_s^* = r_s$, the above equality is trivial. Otherwise, r_s^* is obtained from r_s through a composition of two permutations. The first one permutes lottery numbers only among students in some set I_1 such that $\rho^{\varphi}_{i's} = \rho^{\varphi}_{i''s}$ for all $i', i'' \in I_1$ and there exists i with $D_{is}(r) = 1$ such that $\max_{i' \in I_1}\{\rho^{\varphi}_{i's} + r_{i's}\} \leq \rho^{\varphi}_{is} + r_{is}$. The second permutation permutes lottery numbers only among students in some set I_0 such that $\rho^{\varphi}_{i's} = \rho^{\varphi}_{i''s}$ for all $i', i'' \in I_0$ and $\min_{i' \in I_0}\{\rho^{\varphi}_{i's} + r_{i's}\} > \rho^{\varphi}_{is} + r_{is}$ for all i with $D_{is}(r) = 1$. The first permutation is a composition of special permutations δ satisfying the conditions in Lemma 4. If φ is not strategy-proof for schools, then the proof of Theorem 1 is complete. If φ is strategy-proof for schools, the second permutation is a composition of special permutations δ satisfying the conditions in Lemma 3. Therefore Lemmas 3 and 4 imply Step 2.A. □

Step 6.B. For all lottery number profile r, $\varphi(r_s^*, r_{-s}) = \varphi(r_s^*, \tilde{\sigma}_{-s}(r))$.

Proof of Step 6.B. If $i_0(r) = i_1(r) = \emptyset$ and so $\tilde{\sigma}_{-s}(r) = r_{-s}$, the above inequality is trivial. Otherwise, at the first step of the DA algorithm constituting φ, students apply for schools in the same way both under (r_s^*, r_{-s}) and $(r_s^*, \tilde{\sigma}_{-s}(r))$. In particular, $i_1(r)$ applies for s since $i_1(r) \in First_s(r)$. Schools also tentatively accept students in the same way both under (r_s^*, r_{-s}) and $(r_s^*, \tilde{\sigma}_{-s}(r))$: s does so since s has the same strict priority $\succ^{\varphi}_{r_s^*}$ both under (r_s^*, r_{-s}) and $(r_s^*, \tilde{\sigma}_{-s}(r))$. The other schools also do so since the only possible difference between $\succ^{\varphi}_{r_{s'}}$ and $\succ^{\varphi}_{\tilde{\sigma}_{s'}(r)}$ is the position of $i_1(r)$, who applies for s. As a result, since Step 2.A implies $D_{i_1(r)s}(r_s^*, r_{-s}) = 1$, s tentatively accepts $i_1(r)$ at the first step of the DA algorithm both under (r_s^*, r_{-s}) and $(r_s^*, \tilde{\sigma}_{-s}(r))$. Since (a) s has the same preference $\succ^{\varphi}_{r_s^*}$ both under (r_s^*, r_{-s}) and $(r_s^*, \tilde{\sigma}_{-s}(r))$, (b) the only possible difference between $\succ^{\varphi}_{r_{s'}}$ and $\succ^{\varphi}_{\tilde{\sigma}_{s'}(r)}$ is the position of $i_1(r)$, and (c) $i_1(r)$ is tentatively kept by s and is never be rejected by s under (r_s^*, r_{-s}), the DA algorithm operates in the same way for the remaining steps, producing the same matching. □

Lemma 7. *(Partial outcome-equivalence between $\sigma^*(r)$ and $(\sigma_s^*(r_s^*, r_{-s}), \tilde{\sigma}_{-s}(r))$) For all lottery number profile r with $D_{i_0(r)s}\,(\sigma^*(r)) = D_{i_1(r)s}(\sigma^*(r)) = 0$, it is the case that $D_{i_0(r)s}(\sigma_s^*(r_s^*, r_{-s}), \tilde{\sigma}_{-s}(r)) = D_{i_1(r)s}(\sigma_s^*(r_s^*, r_{-s}), \tilde{\sigma}_{-s}(r)) = 0$.*

Proof of Lemma 7. If $i_0(r) = i_1(r) = \emptyset$ and so $\sigma^*(r) = r = (\sigma_s^*(r_s^*, r_{-s}), \tilde{\sigma}_{-s}(r))$, Lemma 7 is immediate. Otherwise, I first prove the following result.

Step 7.A. For all lottery number profile r with $D_{i_0(r)s}(\sigma^*(r)) = D_{i_1(r)s}(\sigma^*(r)) = 0$, it is the case $\varphi(\sigma^*(r)) = \varphi(\sigma_s^*(r_s^*, r_{-s}), \sigma_{-s}^*(r))$.

Proof of Step 7.A. Note that

a) $\sigma^*_{i_0(r)s}(r) = r_{i_1(r)s} \leq \min\{r_{i_0(r)s}, r^*_{i_0(r)s}, r^*_{i_1(r)s}\} = \min\{\sigma^*_{i_1(r)s}(r), \sigma^*_{i_1(r)s}(r^*_s, r_{-s}), \sigma^*_{i_0(r)s}(r^*_s, r_{-s})\}$, where the last equality follows from $i_0(r^*_s, r_{-s}) = i_0(r)$ and $i_1(r^*_s, r_{-s}) = i_1(r)$ by Step 2.A.

b) $\rho^\varphi_{js} + \sigma^*_{js}(r) > \rho^\varphi_{is} + \sigma^*_{is}(r)$ for all j with $\rho^\varphi_{js} = \rho^\varphi_{i_0(r)s}$ and $\sigma^*_{js}(r) > \sigma^*_{i_0(r)s}(r)$ and all i with $D_{is}(\sigma^*(r)) = 1$ since $D_{i_0(r)s}(\sigma^*(r)) = 0$ by assumption and $i_0(r) \in First_s(r)$ and so $i_0(r)$ ranks s first.

(a) and (b) imply that starting from $\sigma^*(r)$, $\sigma^*_s(r^*_s, r_{-s})$ is obtained from $\sigma^*_s(r)$ through a permutation that permutes lottery numbers only among students in some set I_0 such that $\rho^\varphi_{i's} = \rho^\varphi_{i''s}$ for all $i', i'' \in I_0$ and $\min_{i' \in I_0}\{\rho^\varphi_{i's} + \sigma^*_{i's}(r)\} > \rho^\varphi_{is} + \sigma^*_{is}(r)$ for all i with $D_{is}(\sigma^*(r)) = 1$. This permutation is a composition of special permutations δ's that satisfy the conditions in Lemma 3. Therefore Lemmas 3 and 4 imply Step 7.A. $\square$

Now let me compare $\varphi(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_{-s}(r))$ and $\varphi(\sigma^*_s(r^*_s, r_{-s}), \tilde{\sigma}_{-s}(r))$. At the first step of the DA algorithm constituting φ, students apply for schools in the same way both under $(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_{-s}(r))$ and $(\sigma^*_s(r^*_s, r_{-s}), \tilde{\sigma}_{-s}(r))$. In particular, $i_0(r)$ and $i_1(r)$ apply for s since $i_0(r), i_1(r) \in First_s(r)$. Schools also tentatively accept students in the same way both under $(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_{-s}(r))$ and $(\sigma^*_s(r^*_s, r_{-s}), \tilde{\sigma}_{-s}(r))$: s does so since s has the same strict priority $\succ^\varphi_{\sigma^*_s(r^*_s, r_{-s})}$ both under $(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_{-s}(r))$ and $(\sigma^*_s(r^*_s, r_{-s}), \tilde{\sigma}_{-s}(r))$. The other schools also do so since the only possible differences between $\succ^\varphi_{\sigma^*_{s'}(r)}$ and $\succ^\varphi_{\tilde{\sigma}_{s'}(r)}$ are the positions of $i_0(r)$ and $i_1(r)$, both of whom apply for s.

If s rejects both $i_0(r)$ and $i_1(r)$ at the first step, the proof is complete. Otherwise, s tentatively accepts at least $i_0(r)$ since $\rho^\varphi_{i_0(r)s} + \sigma^*_{i_0(r)s}(r^*_s, r_{-s}) < \rho^\varphi_{i_1(r)s} + \sigma^*_{i_1(r)s}(r^*_s, r_{-s})$. Since $\succ^\varphi_{\sigma^*_{s'}(r)}$ and $\succ^\varphi_{\tilde{\sigma}_{s'}(r)}$ are equivalent over $I \setminus \{i_0(r)\}$, the remaining steps of the DA algorithm operate in the same way both under $(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_{-s}(r))$ and $(\sigma^*_s(r^*_s, r_{-s}), \tilde{\sigma}_{-s}(r))$ until the point where s rejects $i_0(r)$. s finally rejects $i_0(r)$ since it does so under $(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_{-s}(r))$ by Step 7.A and s has the same preference $\succ^\varphi_{\sigma^*_s(r^*_s, r_{-s})}$ both under $(\sigma^*_s(r^*_s, r_{-s}), \sigma^*_{-s}(r))$ and $(\sigma^*_s(r^*_s, r_{-s}), \tilde{\sigma}_{-s}(r))$. This implies Lemma 7. $\square$

Lemma 8. *(Existence of a profitable preference manipulation) There exist $(\rho^*, r^*) \in \Gamma^{|S|}$, school s's preference $\succ_s$ responsive with respect to (c_s, ρ^*_s, r^*_s), and $(\rho'_s, r'_s) \in \Gamma$ such that $\varphi_s((\rho'_s, r'_s), (\rho^*_{-s}, r^*_{-s})) \succ_s \varphi_s(\rho^*, r^*)$ where $(\rho^*_{-s}, r^*_{-s}) \equiv (\rho^*_{s'}, r^*_{s'})_{s' \neq s}$.*

Proof of Lemma 8. Lemmas 5, 6, and 7 imply that there exists r such that

- $D_{i_0(r)s}(r^*_s, \tilde{\sigma}_{-s}(r)) = 0$

- $D_{i_1(r)s}(r_s^*, \tilde{\sigma}_{-s}(r)) = 1$
- $D_{i_0(r)s}(\sigma_s^*(r_s^*, r_{-s}), \tilde{\sigma}_{-s}(r)) = D_{i_1(r)s}(\sigma_s^*(r_s^*, r_{-s}), \tilde{\sigma}_{-s}(r)) = 0.$

*Step 8.A. $\varphi_s(r_s^*, \tilde{\sigma}_{-s}(r)) = \mu_{st}(r_s^*, \tilde{\sigma}_{-s}(r))$ where $\mu_{st}(\cdot)$ is s's tentative assignment at the end of step t in **2STAGES**(r).*

Proof of Step 8.A. Execute `STAGE1`$(r_s^*, \tilde{\sigma}_{-s}(r))$ and start `STAGE2`$(r_s^*, \tilde{\sigma}_{-s}(r))$. s rejects $i_0(r)$ and tentatively keeps $i_1(r)$ since $D_{i_0(r)s}(r_s^*, \tilde{\sigma}_{-s}(r)) = 0$ and $D_{i_1(r)s}(r_s^*, \tilde{\sigma}_{-s}(r)) = 1$. Suppose to the contrary $\varphi_s(r_s^*, \tilde{\sigma}_{-s}(r)) \neq \mu_{st}(r_s^*, \tilde{\sigma}_{-s}(r))$. Since $|\varphi_s(r_s^*, \tilde{\sigma}_{-s}(r))| = |\mu_{st}(r_s^*, \tilde{\sigma}_{-s}(r))| = c_s$ (because s rejects $i_0(r)$ when choosing $\mu_{st}(r_s^*, \tilde{\sigma}_{-s}(r))$ at step t), this implies there exists a student $i_2 \in \varphi_s(r_s^*, \tilde{\sigma}_{-s}(r))$ such that $i_2 \notin \mu_{st}(r_s^*, \tilde{\sigma}_{-s}(r))$. In addition, $i_2 \notin \mu_{st-1}(r_s^*, \tilde{\sigma}_{-s}(r)) \cup i_1(r)$ has to be the case since otherwise (i.e., if $i_2 \notin \mu_{st}(r_s^*, \tilde{\sigma}_{-s}(r))$ but $i_2 \in \mu_{st-1}(r_s^*, \tilde{\sigma}_{-s}(r)) \cup i_1(r)$ and so i_2 applies for s at step $t' < t$ in `STAGE1`$(r_s^*, \tilde{\sigma}_{-s}(r))$), s rejects i_2 at step t and so $i_2 \notin \varphi_s(r_s^*, \tilde{\sigma}_{-s}(r))$, a contradiction. This means i_2 applies for s at a step $t' > t$ and is tentatively kept by s. This requires that s rejects $i_1(r)$ before or at step t' since by definition $i \succ^{\varphi}_{r_s^*} i_1(r)$ for any $i \in \mu_{st}(r_s^*, \tilde{\sigma}_{-s}(r)) \setminus i_1(r)$, which is because s rejects $i_0(r)$ at step t and there is no such i that $\rho^{\varphi}_{is} = \rho^{\varphi}_{i_0(r)s} = \rho^{\varphi}_{i_1(r)s}$ and $r^*_{i_0(r)s} > r^*_{is} > r^*_{i_1(r)s}$. This is a contradiction to the above fact that $D_{i_1(r)s}(r_s^*, \tilde{\sigma}_{-s}(r)) = 1$. □

Step 8.B. There exists a step $t' > t$ in `2STAGES`$(\sigma_s^*(r_s^*, r_{-s}), \tilde{\sigma}_{-s}(r))$ *such that $\mu_{st'}(\sigma_s^*(r_s^*, r_{-s}), \tilde{\sigma}_{-s}(r)) = \mu_{st}(r_s^*, \tilde{\sigma}_{-s}(r)) \cup i_2 \setminus i_1(r)$ where i_2 is a student with $i_2 \succ^{\varphi}_{r_s^*} i_1(r)$.*

Proof of Step 8.B. Execute `STAGE1`$(\sigma_s^*(r_s^*, r_{-s}), \tilde{\sigma}_{-s}(r))$ and start `STAGE2`$(\sigma_s^*(r_s^*, r_{-s}), \tilde{\sigma}_{-s}(r))$. School s rejects $i_1(r)$ and tentatively keeps $i_0(r)$ at step t since (1) μ_{st-1} is the same between $(r_s^*, \tilde{\sigma}_{-s}(r))$ and $(\sigma_s^*(r_s^*, r_{-s}), \tilde{\sigma}_{-s}(r))$, (2) $\rho^{\varphi}_{i_1(r)s} + \sigma^*_{i_1(r)s}(r_s^*, r_{-s}) = \rho^{\varphi}_{i_0(r)s} + r^*_{i_0(r)s}$, and (3) $D_{i_1(r)s}(r_s^*, \tilde{\sigma}_{-s}(r)) = 1$. By $D_{i_0(r)s}(\sigma_s^*(r_s^*, r_{-s}), \tilde{\sigma}_{-s}(r)) = 0$, school s rejects $i_0(r)$ at a later step $t' > t$ and tentatively keeps i_2 with $i_2 \succ^{\varphi}_{\sigma_s^*(r_s^*, r_{-s})} i_0(r)$, which implies $i_2 \succ^{\varphi}_{r_s^*} i_1(r)$ since $\rho^{\varphi}_{i_0(r)s} + \sigma^*_{i_0(r)s}(r_s^*, r_{-s}) = \rho^{\varphi}_{i_1(r)s} + r^*_{i_1(r)s}$ and $\sigma^*_{i_2 s}(r_s^*, r_{-s}) = r^*_{i_2 s}$. Therefore, $\mu_{st'}(\sigma_s^*(r_s^*, r_{-s}), \tilde{\sigma}_{-s}(r)) = \mu_{st}(r_s^*, \tilde{\sigma}_{-s}(r)) \cup i_2 \setminus i_1(r)$ where i_2 is a student with $i_2 \succ^{\varphi}_{r_s^*} i_1(r)$. □

I am ready to construct a profitable preference manipulation for s. Let $\succ_s$ be any preference for s that is responsive with respect to (c_s, ρ_s, r_s^*). Let ρ'_s be a coarse priority order for s such that $\rho'_{ks} > \rho'_{js}$ for all $k \notin \cup_{t_0=1}^{t} \mu_{st_0}(r_s^*, \tilde{\sigma}_{-s}(r)) \cup i_2 \setminus i_1(r)$ and $j \in \cup_{t_0=1}^{t} \mu_{st_0}(r_s^*, \tilde{\sigma}_{-s}(r)) \cup i_2 \setminus i_1(r)$ while $\rho'_{ks} = \rho'_{js}$ if and only if $\rho_{ks} = \rho_{js}$ for all $j, k \notin \mu_{st}(r_s^*, \tilde{\sigma}_{-s}(r)) \cup i_2 \setminus i_1(r)$ or $j, k \in \mu_{st}(r_s^*, \tilde{\sigma}_{-s}(r)) \cup i_2 \setminus i_1(r)$. By Steps 8.A and 8.B, $\varphi_s(r_s^*, \tilde{\sigma}_{-s}(r)) = \mu_{st}(r_s^*, \tilde{\sigma}_{-s}(r))$ while $\varphi_s((\rho'_s, \rho_{-s}), (\sigma_s^*(r_s^*, r_{-s}), \tilde{\sigma}_{-s}(r)) = \mu_{st}(r_s^*, \tilde{\sigma}_{-s}(r)) \cup i_2 \setminus i_1(r)$.

Also, $i_2 \succ^{\varphi}_{r^*_s} i_1(r)$ established in Step 8.B implies $i_2 \succ_{r^*_s} i_1(r)$ as follows:

$$
\begin{aligned}
& i_2 \succ^{\varphi}_{r^*_s} i_1(r) \\
\Leftrightarrow\ & f^{\varphi}(\rho_{i_2 s}) + g^{\varphi}(rank_{i_2 s}) + r^*_{i_2 s} < f^{\varphi}(\rho_{i_1(r)s}) + g^{\varphi}(rank_{i_1(r)s}) + r^*_{i_1(r)s} \\
\Rightarrow\ & f^{\varphi}(\rho_{i_2 s}) + r^*_{i_2 s} < f^{\varphi}(\rho_{i_1(r)s}) + r^*_{i_1(r)s} \\
& \text{(since } rank_{i_1(r)s} = 1 \leq rank_{i_2 s} \text{ and } g^{\varphi}(\cdot) \text{ is weakly increasing)} \\
\Leftrightarrow\ & \rho_{i_2 s} + r^*_{i_2 s} < \rho_{i_1(r)s} + r^*_{i_1(r)s} \\
& \text{(since } f^{\varphi}(\cdot) \text{ is strictly increasing)} \\
\Leftrightarrow\ & i_2 \succ_{r^*_s} i_1(r).
\end{aligned}
$$

Thus $\varphi_s((\rho'_s, \rho_{-s}), (\sigma^*_s(r^*_s, r_{-s}), \tilde{\sigma}_{-s}(r)) = \mu_{st}(r^*_s, \tilde{\sigma}_{-s}(r)) \cup i_2 \setminus i_1(r) \succ_s \mu_{st}(r^*_s, \tilde{\sigma}_{-s}(r)) = \varphi_s(\rho, (r^*_s, \tilde{\sigma}_{-s}(r)))$ since $\succ_s$ is responsive with respect to (c_s, ρ_s, r^*_s), showing that when (ρ_s, r^*_s) is s's true private information, $(\rho'_s, \sigma^*_s(r^*_s, r_{-s}))$ is a profitable manipulation for s with respect to any $\succ_s$ responsive with respect to (c_s, ρ_s, r^*_s); therefore gDA mechanism φ is not strategy-proof for schools. Figure 3 summarizes the structure of the above proof for Case 1. □

Figure 3: Structure of the proof (Case 1)

The first-choice research design does not extracts a random assignment for school s for gDA mechanism φ

$\Downarrow$

$$
\begin{tikzcd}
(\rho, r) \arrow[rr, "\text{A breakdown of the Fisher property}", "\text{(Lemma 5)}"'] \arrow[dd, squiggly, no head, "\text{Outcome-equivalent}"', "\text{(Lemma 6)}"] & & (\rho, \sigma^*(r)) \arrow[d, squiggly, no head, "\text{Partially outcome-equivalent}"', "\text{(Lemma 7)}"] \\
& & (\rho, (\sigma^*_s(r^*_s, r_{-s}), \tilde{\sigma}_{-s}(r))) \arrow[d, "\text{School } s \text{ relegates ``nuisance students''}"] \\
(\rho, (r^*_s, \tilde{\sigma}_{-s}(r))) \arrow[rr, "\text{Profitable manipulation}", "\text{for school } s \text{ (Lemma 8)}"'] & & ((\rho'_s, \rho_{-s}), (\sigma^*_s(r^*_s, r_{-s}), \tilde{\sigma}_{-s}(r)))
\end{tikzcd}
$$

$\Downarrow$

gDA mechanism φ is not strategy-proof for school s

<u>Case 2</u>: *There exist lottery number profile r consistent with φ's lottery structure (STB or MTB), student $i \in First_s(r)$, and permutation $\hat{\sigma}(r)$ defined right before Lemma 5 such that*

$D_{is}(\hat{\sigma}(r)) \neq D_{is}(r)$. Recall that $\hat{\sigma}(r)$ is a permutation that is different from $\sigma^*_s(r)$ and switches only two students i' and i'' who are consecutive in r_s within $First_s(r)$. Since $\hat{\sigma}(r)$ is defined to be different from $\sigma^*_s(r)$, it is the case $D_{i's}(r) = D_{i''s}(r)$. Depending on whether $D_{i's}(r) = D_{i''s}(r) = 0$ or 1, there are two cases to consider.

<u>*Case 2.a*</u>: $D_{i's}(r) = D_{i''s}(r) = 1$. Then there exists j with $D_{is}(r) = 1$ such that $\max\{\rho^{\varphi}_{i's} + r_{i's}, \rho^{\varphi}_{i''s} + r_{i''s}\} \leq \rho^{\varphi}_{js} + r_{js}$. If φ uses MTB, $\hat{\sigma}(r)$ satisfies the conditions in Lemma 4, implying $\varphi(\hat{\sigma}(r)) = \varphi(r)$ by Lemma 4. This is a contradiction to $D_{is}(\hat{\sigma}(r)) \neq D_{is}(r)$.

If φ uses STB, suppose to the contrary that $D_{i's}(r) = D_{i''s}(r) = 1$. At the first step of the DA algorithm, students apply for schools in the same way both under r and $\hat{\sigma}(r)$. In particular, i' and i'' apply for s since $i', i'' \in First_s(r)$ and both of them rank s first. Schools also tentatively accept students in the same way both under r and $\hat{\sigma}(r)$: The other schools than s do so since the only possible differences between $\succ^{\varphi}_{r_{s'}}$ and $\succ^{\varphi}_{\hat{\sigma}_{s'}(r)}$ are the positions of i' and i'', both of whom apply for s. s accepts the same students including i' and i'' since $D_{i's}(r) = D_{i''s}(r) = 1$ and $\{\rho^{\varphi}_{js} + r_{js} | j$ applies for s at the first step of the DA algorithm under $r\} = \{\rho^{\varphi}_{js} + \hat{\sigma}_{js}(r) | j$ applies for s at the first step of the DA algorithm under $\hat{\sigma}(r)\}$, which is because the same students apply for s both under r and $\hat{\sigma}(r)$, $\rho^{\varphi}_{js} + r_{js} = \rho^{\varphi}_{js} + \hat{\sigma}_{js}(r)$ for all $j \neq i', i''$, $\rho^{\varphi}_{i's} + r_{i's} = \rho^{\varphi}_{i''s} + \hat{\sigma}_{i''s}(r)$, and $\rho^{\varphi}_{i''s} + r_{i''s} = \rho^{\varphi}_{i's} + \hat{\sigma}_{i's}(r)$. Since (a) the only possible differences between $\succ^{\varphi}_{r_{s'}}$ and $\succ^{\varphi}_{\hat{\sigma}_{s'}(r)}$ are the positions of i' and i'', and (b) i' and i'' are tentatively kept by s and is never be rejected by s under r, the DA algorithm operates in the same way for the remaining steps, producing the same matching. This implies $D_{is}(\hat{\sigma}(r)) = D_{is}(r)$ for all $i \in First_s(r)$, a contradiction.

<u>*Case 2.b*</u>: $D_{i's}(r) = D_{i''s}(r) = 0$. Without loss of generality, assume that there exist r and $i \in First_s(r)$ such that $D_{is}(\hat{\sigma}(r)) = 0 \neq 1 = D_{is}(r)$. Let i^* be the student with $i^* \in First_s(r)$, $D_{i^*s}(r) = 1$, and $r_{i^*s} \geq r_{js}$ for all $j \in First_s(r)$ with $D_{js}(r) = 1$. Until the end of Case 2.b, change i^*'s preference $\succ_{i^*}$ to $\succ'_{i^*}$ such that $s \succ'_{i^*} \emptyset \succ'_{i^*} s'$ for all $s' \neq s$. This does not change $D_{i^*s}(r) = 1$ or $D_{i's}(r) = D_{i''s}(r) = 0$. Note that $D_{i^*s}(\hat{\sigma}(r)) = 0 \neq 1 = D_{i^*s}(r)$ since $D_{is}(\hat{\sigma}(r)) = 0$ and $\rho^{\varphi}_{i^*s} + \hat{\sigma}_{i^*s}(r) = \rho^{\varphi}_{i^*s} + r_{i^*s} \geq \rho^{\varphi}_{is} + r_{is} = \rho^{\varphi}_{is} + \hat{\sigma}_{is}(r)$ for any $i \in First_s(r)$ with $D_{is}(\hat{\sigma}(r)) = 0 \neq 1 = D_{is}(r)$. Without loss of generality, assume $r_{i's} < r_{i''s}$ so that $\hat{\sigma}_{i''s}(r) < \hat{\sigma}_{i's}(r)$. Let $\hat{\sigma}^{\#}_s(r)$ be the further permutation of $\hat{\sigma}_s(r)$ such that $\hat{\sigma}^{\#}_{i''s}(r) = \min\{\hat{\sigma}_{js}(r) | j \in First_s(r), D_{js}(r) = 0\}$, $\hat{\sigma}^{\#}_{i's}(r) = \min\{\hat{\sigma}_{js}(r) \neq \hat{\sigma}^{\#}_{i''s}(r) | j \in First_s(r), D_{js}(r) = 0\}$, and $\hat{\sigma}^{\#}_{js}(r) > \hat{\sigma}^{\#}_{ks}(r)$ if and only if $\hat{\sigma}_{js}(r) > \hat{\sigma}_{ks}(r)$ for all $j, k \in I \setminus \{i', i''\}$.

Lemma 9. *$D_{i^*s}(\hat{\sigma}^{\#}_s(r), \hat{\sigma}_{-s}(r)) = D_{i''s}(\hat{\sigma}^{\#}_s(r), \hat{\sigma}_{-s}(r)) = 0$ where $\hat{\sigma}_{-s}(r) \equiv (\hat{\sigma}_{s'}(r))_{s' \in S}$.*

Proof of Lemma 9. Note that $\min\{\rho^{\varphi}_{i's} + \hat{\sigma}_{i's}(r), \rho^{\varphi}_{i''s} + \hat{\sigma}_{i''s}(r)\} = \min\{\rho^{\varphi}_{i's} + r_{i's}, \rho^{\varphi}_{i''s} + r_{i''s}\} >$

$\rho^{\varphi}_{i^*s} + r_{i^*s} = \rho^{\varphi}_{i^*s} + \hat{\sigma}_{i^*s}(r)$, where the first and last equalities are by $\rho^{\varphi}_{i's} = \rho^{\varphi}_{i''s}$ (implied by $i', i'' \in First_s(r)$) and the definition of $\hat{\sigma}_s(r)$ while the middle inequality is by $D_{i^*s}(r) = 1$ and $D_{i's}(r) = D_{i''s}(r) = 0$. Thus, $D_{i''s}(\hat{\sigma}(r)) = D_{i's}(\hat{\sigma}(r)) = 0$. If φ is not strategy-proof for schools, then the proof of Theorem 1 is complete. If φ is strategy-proof for schools, by Lemma 3 and the definition of $\hat{\sigma}^{\#}_s(r)$, it holds $\varphi(\hat{\sigma}^{\#}_s(r), \hat{\sigma}_{-s}(r)) = \varphi(\hat{\sigma}(r))$, which implies Lemma 9. □

Let $\hat{\sigma}^{\#\#}_s(r)$ be the permutation of $\hat{\sigma}^{\#}_s(r)$ that switches i^* and i'', who are consecutive within $First_s(r)$ under $\hat{\sigma}^{\#}_s(r)$.

Lemma 10. *$D_{i^*s}(\hat{\sigma}^{\#\#}_s(r), \hat{\sigma}_{-s}(r)) = 0$ and $D_{i''s}(\hat{\sigma}^{\#\#}_s(r), \hat{\sigma}_{-s}(r)) = 1$.*

Proof of Lemma 10. Note that $D_{i^*s}(\hat{\sigma}^{\#}_s(r), r_{-s}) = 1$ and $D_{i''s}(\hat{\sigma}^{\#}_s(r), r_{-s}) = 0$ by Lemma 3 and the definition of $\hat{\sigma}^{\#}_s(r)$. Since the only differences between $(\hat{\sigma}^{\#}_s(r), r_{-s})$ and $(\hat{\sigma}^{\#\#}_s(r), \hat{\sigma}_{-s}(r))$ are the positions of i^* and i'' in the priority order at s and the positions of i' and i'' in the priority order at $s' \neq s$, both under $(\hat{\sigma}^{\#}_s(r), r_{-s})$ and $(\hat{\sigma}^{\#\#}_s(r), \hat{\sigma}_{-s}(r))$, the DA algorithm operates in the same way until i' is rejected by s. School s rejects i' in both scenarios since $D_{i''s}(\hat{\sigma}^{\#}_s(r), \hat{\sigma}_{-s}(r)) = 0$ (as shown at the start of this proof) and $\rho^{\varphi}_{i's} + \hat{\sigma}^{\#\#}_{i's}(r) = \rho^{\varphi}_{i's} + \hat{\sigma}^{\#}_{i's}(r) > \rho^{\varphi}_{i''s} + \hat{\sigma}^{\#}_{i''s}(r) = \rho^{\varphi}_{i^*s} + \hat{\sigma}^{\#\#}_{i^*s}(r)$.

Since i' has a weakly worse lottery number under $\hat{\sigma}_{s'}(r)$ than under $r_{s'}$ for all $s' \neq s$, i is less likely to crowd other applicants out from other schools than s and the chain reactions of new rejections and applications caused by s's rejection of i' are less likely to go back to s under $(\hat{\sigma}^{\#\#}_s(r), \hat{\sigma}_{-s}(r))$ than under $(\hat{\sigma}^{\#}_s(r), r_{-s})$. Also, since the only other difference between $(\hat{\sigma}^{\#\#}_s(r), \hat{\sigma}_{-s}(r))$ and $(\hat{\sigma}^{\#}_s(r), r_{-s})$ is the school-s lottery numbers of i'' and i^* , i.e., $\hat{\sigma}^{\#}_{i^*s}(r) = \hat{\sigma}^{\#\#}_{i''s}(r) \neq \hat{\sigma}^{\#}_{i''s}(r) = \hat{\sigma}^{\#\#}_{i^*s}(r)$, and $\rho^{\varphi}_{i''s} = \rho^{\varphi}_{i^*s}$, when i'' is rejected by s under $(\hat{\sigma}^{\#}_s(r), r_{-s})$, i^* may be rejected by s under $(\hat{\sigma}^{\#\#}_s(r), \hat{\sigma}_{-s}(r))$. But i^* ranks only s in $\succ'_{i^*}$ and causes no additional rejections at other schools while i'' may rank other schools than s and may cause additional rejections at other schools.

By these two factors, the set of rejections made by schools other than s is weakly larger in the set inclusion sense under $(\hat{\sigma}^{\#}_s(r), r_{-s})$ than under $(\hat{\sigma}^{\#\#}_s(r), \hat{\sigma}_{-s}(r))$, i.e., $\{(j, s') | D_{js''}(\hat{\sigma}^{\#\#}_s (r), \hat{\sigma}_{-s}(r)) = 0$ for all $s'' \succeq_j s'\} \subseteq \{(j, s') | D_{js''}(\hat{\sigma}^{\#}_s(r), r_{-s}) = 0$ for all $s'' \succeq_j s'\}$. Therefore, the set of applicants for s is weakly larger in the set inclusion sense under $(\hat{\sigma}^{\#}_s(r), r_{-s})$ than under $(\hat{\sigma}^{\#\#}_s(r), \hat{\sigma}_{-s}(r))$, i.e., $\{j | D_{js'}(\hat{\sigma}^{\#\#}_s(r), \hat{\sigma}_{-s}(r)) = 0$ for all $s' \succ_j s\} \subseteq \{j | D_{js'}(\hat{\sigma}^{\#}_s(r), r_{-s}) = 0$ for all $s' \succ_j s\}$. As a result it has to be the case that the cutoff at s is smaller (more strict) under $(\hat{\sigma}^{\#}_s(r), r_{-s})$ than under $(\hat{\sigma}^{\#\#}_s(r), \hat{\sigma}_{-s}(r))$, i.e.,

$$\max\{\rho^{\varphi}_{js} + \hat{\sigma}^{\#\#}_{js}(r) | D_{js}(\hat{\sigma}^{\#\#}_s(r), \hat{\sigma}_{-s}(r)) = 1\}$$

$$
\begin{aligned}
&= c_s\text{-th}\{\rho^{\varphi}_{js} + \hat{\sigma}^{\#\#}_{js}(r) | D_{js'}(\hat{\sigma}^{\#\#}_{s}(r), \hat{\sigma}_{-s}(r)) = 0 \text{ for all } s' \succ_j s\} \\
&\geq c_s\text{-th}\{\rho^{\varphi}_{js} + \hat{\sigma}^{\#}_{js}(r) | D_{js'}(\hat{\sigma}^{\#}_{s}(r), r_{-s}) = 0 \text{ for all } s' \succ_j s\} \\
&= \max\{\rho^{\varphi}_{js} + \hat{\sigma}^{\#}_{js}(r) | D_{js}(\hat{\sigma}^{\#}_{s}(r), r_{-s}) = 1\} \\
&\geq \rho^{\varphi}_{i^*s} + r_{i^*s} \\
&= \rho^{\varphi}_{i''s} + \hat{\sigma}^{\#\#}_{i''s}(r),
\end{aligned}
$$

where the first inequality is by $\{j | D_{js'}(\hat{\sigma}^{\#\#}_{s}(r), \hat{\sigma}_{-s}(r)) = 0 \text{ for all } s' \succ_j s\} \subseteq \{j | D_{js'}(\hat{\sigma}^{\#}_{s}(r), r_{-s}) = 0 \text{ for all } s' \succ_j s\}$ (shown above), the second inequality is by $D_{i^*s}(\hat{\sigma}^{\#}_{s}(r), r_{-s}) = 1$ (shown at the start of this proof), and the last equality is by the definition of $\hat{\sigma}^{\#\#}_{i''s}(r)$. c_s-th$\{\cdot\}$ is the c_s-th order statistic. This implies $D_{i''s}(\hat{\sigma}^{\#\#}_{s}(r), \hat{\sigma}_{-s}(r)) = 1$ (by $i'' \in First_s(r)$). This also implies $D_{i^*s}(\hat{\sigma}^{\#\#}_{s}(r), \hat{\sigma}_{-s}(r)) = 0$ since otherwise $D_{i^*s}(\hat{\sigma}^{\#}_{s}(r), \hat{\sigma}_{-s}(r)) = D_{i''s}(\hat{\sigma}^{\#}_{s}(r), \hat{\sigma}_{-s}(r)) = 1$ by Lemma 4, a contradiction to $D_{i^*s}(\hat{\sigma}^{\#}_{s}(r), \hat{\sigma}_{-s}(r)) = 0$ in Lemma 9. This completes the proof of Lemma 10. □

Lemmas 9 and 10 imply gDA mechanism φ is not strategy-proof for schools by the same argument as Case 1 where students i'' and i^* perform the roles of students $i_1(r)$ and $i_0(r)$, respectively, in Case 1 while the permutation from lottery number profile $(\hat{\sigma}^{\#\#}_{s}(r), \hat{\sigma}_{-s}(r))$ to $(\hat{\sigma}^{\#}_{s}(r), \hat{\sigma}_{-s}(r))$ performs the role of the permutation from lottery number profile r to $\sigma^*(r)$ in Case 1.

A.2 Proof of Corollary 1.b

Proof. Consider a special case of the proof of Theorem 1 where I suppose that the first-choice research design does not extract a random assignment for the DA mechanism with STB when there are no priorities, i.e., $\rho_{is} = \rho_{js}$ for all i, j, and s. By the STB lottery structure, $r_{is'} = r_{is''}$ for all i, s', and s'', and the order of r_{is} is the same as the order of ρ^{φ}_{is} for any s. In this case, Case 2 never happens and only Case 1 is relevant. In Case 1, by the no-priority and STB assumptions, $r^*_s = \tilde{\sigma}_{s'}(r)$ for all $s' \neq s$. This implies that under the preferences induced by $(\rho, (r^*_s, \tilde{\sigma}_{-s}(r)))$, all schools share the same preference as s's $\succ_{r^*_s}$. The proof of Theorem 1 implies that when all the other other schools than s commonly report (ρ_s, r^*_s), reporting $(\rho'_s, \sigma^*_s(r^*_s, r_{-s}))$ is a profitable preference manipulation for s with respect to $\succ_{r^*_s}$. This contradicts the fact that for the DA mechanism, truth-telling is optimal for s when all the other schools report the same preference as s's true preference. (For a formal proof of this well-known fact, see Hatfield et al. (2016) Proposition 4 and Lemma 1.) Therefore, for problems with no priorities, the first-choice research design must extract a random assignment for the DA mechanism with STB. □

A.3 Proof of Proposition 2

Proof. Take any three schools and label them as A, B, and C. Consider any student preference profile that can be written as follows for some $k \neq l$ and $k, l \geq 1$:

$\succ_1$: $B, A, \emptyset$
$\succ_2, ..., \succ_k$: $B, \emptyset$
$\succ_{k+1}$: $C, A, \emptyset$
$\succ_{k+2}, ..., \succ_{k+l}$: $C, \emptyset$
$\rho_A, \rho_B, \rho_C : \{1, 2, 3, 4, 5\}$.

Without loss of generality, assume $l > k$. If $k = 1$, set $\succ_k$: $B, A, \emptyset$. The capacity of each school is 1 while the treatment school is A. Since both students 1 and $k+1$ rank A second and have the same priority at A, for any gDA mechanism φ, we have

$$\rho^{\varphi}_{1A} \equiv f^{\varphi}(\rho_{1A}) + g^{\varphi}(rank_{1A}) = f^{\varphi}(\rho_{k+1,A}) + g^{\varphi}(rank_{k+1,A}) \equiv \rho^{\varphi}_{k+1,A},$$

which I denote by ρ. Nevertheless, it turns out that for any gDA mechanism with any lottery structure,

$$P(Z_{iA}(R) = 1 | \rho^{\varphi}_{iA} = \rho, \theta_i = \theta_1) < P(Z_{iA}(R) = 1 | \rho^{\varphi}_{iA} = \rho, \theta_i = \theta_{k+1}).$$

To see this, note that for any gDA mechanism, student 1 is assigned to B with probability $\frac{1}{k}$ since only students 1 to k rank B, and all of them rank B first so that for all $i, j \in \{1, ..., k\}$,

$$\rho^{\varphi}_{iB} \equiv f^{\varphi}(\rho_{iB}) + g^{\varphi}(rank_{iB}) = f^{\varphi}(\rho_{jB}) + g^{\varphi}(rank_{jB}) \equiv \rho^{\varphi}_{jB}.$$

Likewise, for any gDA mechanism, student $k+1$ is assigned to C with probability $\frac{1}{l}$. Based on these facts, I first analyze any gDA mechanism with STB by considering the following cases.

<u>*Case i*</u>: Neither student 1 nor $k+1$ applies for A, i.e., 1 and $k+1$ are assigned B and C, respectively. This case happens with probability $\frac{1}{k} \times \frac{1}{l}$ since student 1 is assigned to B with probability $\frac{1}{k}$, student $k+1$ is assigned to C with probability $\frac{1}{l}$, and these two events are independent since there is no overlap between applicants for B and those for C so that $\{R_{iB} | \ i \text{ ranks } B\}$ and $\{R_{iC} | \ i \text{ ranks } C\}$ are independent. In this case, no student

applies for A, and A is undersubscribed. Recall that I define $Z_{iA}(r) = 1$ for all i if there is no j with $D_{jA}(r) = 1$. Both students 1 and $k+1$ are therefore qualified for A, i.e., $Z_{1A}(r) = Z_{k+1,A}(r) = 1$.

<u>*Case ii*</u>: Only student $k+1$ applies for A. This case happens with probability $\frac{1}{k} \times \frac{l-1}{l}$. In this case, student $k+1$ is always assigned to A and qualified for A. By $\rho^{\varphi}_{1A} = \rho^{\varphi}_{k+1,A}$ shown above and the fact that student $k+1$ gets the single seat at A, student 1 is qualified for A (i.e., $Z_{1A}(R) = 1$) if and only if student 1 has a better lottery number than student $k+1$ at A (i.e., $R_{1A} < R_{k+1,A}$). Let $U[a,b]$ be a random variable drawn from the uniform distribution over $[a,b]$, $Beta(\alpha,\beta)$ be a random variable drawn from the beta distribution with parameters (α,β), $f(x;\alpha,\beta)$ and $F(x;\alpha,\beta)$ be the pdf and cdf, respectively, of $Beta(\alpha,\beta)$, and $\Gamma(\cdot)$ be the Gamma function. Conditional on Case ii, student 1 has a better lottery number than student $k+1$ at A and so qualified there ($Z_{1A}(R) = 1$) with the following probability.

$$
\begin{aligned}
&\Pr(R_{1A} < R_{k+1,A} | D_{1B}(R) = 1, D_{k+1,C}(R) = 0) \\
&= \Pr(\min\{R_{1B}, ..., R_{kB}\} < U[\min\{R_{k+2,C}, ..., R_{k+l,C}\}, 1]) \\
&= \Pr(\min\{R_{1A}, ..., R_{kA}\} < U[\min\{R_{k+2,A}, ..., R_{k+l,A}\}, 1]) \\
&= \int_0^1 \Pr(Beta(1,k) < U[x,1]) \times f(x;1,l-1)dx \\
&= \int_0^1 \int_x^1 F(y;1,k) \times \frac{1}{1-x} dy f(x;1,l-1)dx \\
&= \int_0^1 \int_x^1 \frac{\Gamma(1+k)}{\Gamma(1)\Gamma(k)} \int_0^y (1-t)^{k-1} dt \frac{1}{1-x} dy \frac{\Gamma(l)}{\Gamma(1)\Gamma(l-1)} (1-x)^{l-2} dx \\
&= \frac{\Gamma(1+k)\Gamma(l)}{\{\Gamma(1)\}^2\Gamma(k)\Gamma(l-1)} \int_0^1 \int_x^1 (-\frac{(1-y)^k - 1}{k}) dy (1-x)^{l-3} dx \\
&= \frac{\Gamma(1+k)\Gamma(l)}{\{\Gamma(1)\}^2\Gamma(k)\Gamma(l-1)} \int_0^1 (-\frac{\frac{(1-x)^{k+1}}{k+1} + x - 1}{k})(1-x)^{l-3} dx \\
&= \frac{\Gamma(1+k)\Gamma(l)}{\{\Gamma(1)\}^2\Gamma(k)\Gamma(l-1)} \times \frac{k+l}{(1+k)(l-1)(k+l-1)} \\
&= \frac{k(k+l)}{\{\Gamma(1)\}^2(1+k)(k+l-1)} \\
&\equiv p_1(k,l),
\end{aligned}
$$

where the second equality is by the STB lottery structure while the first and third equalities use the following facts, respectively:

- If $X \sim U[0,1]$, then the distribution of X conditional on $X \geq x_0$ is $U[x_0, 1]$ where x_0 is any constant on $[0,1]$.
- R_{iA}'s are i.i.d. samples from $U[0,1]$ while the k-th order statistic of n i.i.d. samples from $U[0,1]$ is distributed according to $Beta(k, n+1-k)$ (Casella and Berger (2002), p.230).

<u>*Case iii*</u>: Only student 1 applies for A. This case happens with probability $\frac{k-1}{k} \times \frac{1}{l}$ by the same reason as in Case ii. In this case, 1 is always assigned A and qualified for A. Since $\rho^{\varphi}_{1A} = \rho^{\varphi}_{k+1,A}$ and student 1 gets the single seat at A, student $k+1$ is qualified for A if and only if student $k+1$ has a better lottery number than 1 at A. By the same reasoning as in Case ii, conditional on Case iii, student $k+1$ has a better lottery number than 1 at A and so qualified there ($Z_{k+1,A}(R) = 1$) with probability $p_1(l,k)$.

<u>*Case iv*</u>: Both students 1 and $k+1$ apply for A. This case happens with probability $\frac{k-1}{k} \times \frac{l-1}{l}$ by the same reason as in Cases iii and iv. In this case, again by $\rho^{\varphi}_{1A} = \rho^{\varphi}_{k+1,A}$, only one of students 1 and $k+1$ with a better lottery number is assigned to A and qualified for A. Conditional on Case iv, student 1 has a better lottery number than $k+1$ at A and so qualified there ($Z_{1A}(R) = 1$) with the following probability.

$$
\begin{aligned}
&\underline{p} \\
&\equiv \Pr(R_{1A} < R_{k+1,A} | D_{1B}(R) = 0, D_{k+1,C}(R) = 0) \\
&= \Pr(U[\min\{R_{2,B}, ..., R_{k,B}\}, 1] < U[\min\{R_{k+2,C}, ..., R_{k+l,C}\}, 1]) \\
&= \Pr(U[\min\{R_{2,A}, ..., R_{k,A}\}, 1] < U[\min\{R_{k+2,A}, ..., R_{k+l,A}\}, 1]) \\
&= \Pr(U[Beta(1, k-1), 1] < U[Beta(1, l-1), 1]) \\
&= \int_0^1 \int_0^1 \Pr(U[x,1] < U[y,1]) \times f(x; 1, k-1) \times f(y; 1, l-1) dx dy \\
&= \int_0^1 \int_0^1 \int_y^1 \frac{\max\{t-x, 0\}}{1-x} \frac{1}{1-y} dt \frac{\Gamma(k)}{\Gamma(1)\Gamma(k-1)} (1-x)^{k-2} \times \frac{\Gamma(l)}{\Gamma(1)\Gamma(l-1)} (1-y)^{l-2} dx dy \\
&= \int_0^1 \int_0^1 \int_y^1 \max\{t-x, 0\} dt \frac{\Gamma(k)}{\Gamma(1)\Gamma(k-1)} (1-x)^{k-3} \times \frac{\Gamma(l)}{\Gamma(1)\Gamma(l-1)} (1-y)^{l-3} dx dy,
\end{aligned}
$$

where the third equality uses the STB lottery regime while the fifth equality uses the fact that $(R_{2A},...,R_{kA})$ and $(R_{k+2,A},...,R_{k+l,A})$ are independent. Letting

$$\begin{aligned}
&\bar{p} \\
&\equiv \Pr(R_{1A} > R_{k+1,A} | D_{1B}(R) = 0, D_{k+1,C}(R) = 0) \\
&= \int_0^1 \int_0^1 \int_x^1 \max\{t-y, 0\} dt \frac{\Gamma(k)}{\Gamma(1)\Gamma(k-1)}(1-x)^{k-3} \times \frac{\Gamma(l)}{\Gamma(1)\Gamma(l-1)}(1-y)^{l-3} dxdy,
\end{aligned}$$

we have

$$\begin{aligned}
&\underline{p}/\bar{p} \\
&= \frac{\int_0^1 \int_0^1 \int_y^1 \max\{t-x, 0\} dt (1-x)^{k-3} \times (1-y)^{l-3} dxdy}{\int_0^1 \int_0^1 \int_x^1 \max\{t-y, 0\} dt (1-x)^{k-3} \times (1-y)^{l-3} dxdy} \\
&\leq 1,
\end{aligned}$$

where the last inequality is because of $l > k$. Therefore, since $\underline{p} + \bar{p} = 1$, we have $\underline{p} \leq \frac{1}{2}$ and $\bar{p} \geq \frac{1}{2}$.

To sum up all cases, students 1 and $k+1$'s qualification probabilities at A are different as follows:

$$\begin{aligned}
&\Pr(Z_{iA}(R) = 1 | \rho^{\varphi}_{iA} = \rho, \theta_i = \theta_1) \\
&= \Sigma_{x=i,ii,iii,iv} \Pr(\text{Case } x) \times \Pr(Z_{iA}(R) = 1 | \rho^{\varphi}_{iA} = \rho, \theta_i = \theta_1, \text{Case } x) \\
&= \frac{1}{k} \times \frac{1}{l} + \frac{1}{k} \times \frac{l-1}{l} \times p_1(k,l) + \frac{k-1}{k} \times \frac{1}{l} + \frac{k-1}{k} \times \frac{l-1}{l} \times \underline{p} \\
&= \underbrace{\frac{1}{k} \times \frac{1}{l} + \frac{1}{k} \times \frac{l-1}{l} + \frac{k-1}{k} \times \frac{1}{l}}_{\equiv p_2(k,l)} + \underbrace{\frac{1}{k} \times \frac{l-1}{l} \times (p_1(k,l) - 1)}_{\equiv p_3(k,l)} + \underbrace{\frac{k-1}{k} \times \frac{l-1}{l} \times \underline{p}}_{\equiv p_4(k,l)} \\
&< p_2(k,l) + \underbrace{\frac{k-1}{k} \times \frac{1}{l} \times (p_1(l,k) - 1)}_{> p_3(k,l)} + \underbrace{\frac{k-1}{k} \times \frac{l-1}{l} \times \bar{p}}_{\geq p_4(k,l) \text{ (by } \bar{p} \geq \underline{p})} \\
&= \frac{1}{k} \times \frac{1}{l} + \frac{1}{k} \times \frac{l-1}{l} + \frac{k-1}{k} \times \frac{1}{l} \times p_1(l,k) + \frac{k-1}{k} \times \frac{l-1}{l} \times \bar{p} \\
&= \Sigma_{x=i,ii,iii,iv} \Pr(\text{Case } x) \times \Pr(Z_{iA}(R) = 1 | \rho^{\varphi}_{iA} = \rho, \theta_i = \theta_{k+1}, \text{Case } x)
\end{aligned}$$

$= P(Z_{iA}(R) = 1|\rho^{\varphi}_{iA} = \rho, \theta_i = \theta_{k+1}),$

where the key inequality $\frac{k-1}{k} \times \frac{1}{l} \times (p_1(l,k) - 1) > p_3(k,l)$ comes from the following facts:

- $\frac{1}{k} \times \frac{l-1}{l} > \frac{k-1}{k} \times \frac{1}{l} > 0$ (by $l > k \geq 2$).
- $p_1(k,l) - 1 < p_1(l,k) - 1 < 0$ (the first inequality is by $l > k$ while the second inequality is because both $p_1(k,l)$ and $p_1(k,l)$ are nondegenerate conditional probabilities).

This proves that there is no gDA mechanism with the STB lottery structure for which the qualification IV research design extracts a random assignment.

For any gDA mechanism with MTB, the argument is simplified as follows.

Case *i*: Neither student 1 nor $k+1$ applies for A, i.e., 1 and $k+1$ are assigned B and C, respectively. This case happens with probability $\frac{1}{k} \times \frac{1}{l}$. In this case, no student applies for A, and A is undersubscribed. Both students 1 and $k+1$ are therefore qualified for A.

Case *ii*: Only student $k+1$ applies for A. This case happens with probability $\frac{1}{k} \times \frac{l-1}{l}$. In this case, student $k+1$ is always assigned A and qualified for A. Student 1 is qualified for A if and only if student 1 has a better lottery number than student $k+1$ at A. By $\rho^{\varphi}_{1A} = \rho^{\varphi}_{k+1,A}$, this happens with probability 1/2 by the MTB lottery structure, where R_{1A} and $R_{k+1,A}$ are i.i.d. even conditional on Case *ii* ($D_{1B}(R) = 1$ and $D_{k+1,C}(R) = 0$).

Case *iii*: Only student 1 applies for A. This case happens with probability $\frac{k-1}{k} \times \frac{1}{l}$. In this case, student 1 is always assigned A and qualified for A. Student $k+1$ is qualified for A if and only if student $k+1$ has a better lottery number than 1 at A. By the same reasoning as in Case *ii*, this happens with probability 1/2.

Case *iv*: Both students 1 and $k+1$ apply for A. This case happens with probability $\frac{k-1}{k} \times \frac{l-1}{l}$. In this case, only one of students 1 and $k+1$ with a better lottery number is assigned A and qualified for A. Conditional on Case *iv*, by $\rho^{\varphi}_{1A} = \rho^{\varphi}_{k+1,A}$, student 1 has a better lottery number than $k+1$ at A with probability 1/2.

To sum up all cases, students 1 and $k+1$'s qualification probabilities at A are different as follows:

$\Pr(Z_{iA}(R) = 1|\rho^{\varphi}_{iA} = \rho, \theta_i = \theta_1)$

$$
\begin{aligned}
&= \Sigma_{x=i,ii,iii,iv} \Pr(\text{Case } x) \times \Pr(Z_{iA}(R) = 1 | \rho_{iA}^{\varphi} = \rho, \theta_i = \theta_1, \text{Case } x) \\
&= \frac{1}{k} \times \frac{1}{l} + \frac{1}{k} \times \frac{l-1}{l} \times \frac{1}{2} + \frac{k-1}{k} \times \frac{1}{l} + \frac{k-1}{k} \times \frac{l-1}{l} \times \frac{1}{2} \\
&< \frac{1}{k} \times \frac{1}{l} + \frac{1}{k} \times \frac{l-1}{l} + \frac{k-1}{k} \times \frac{1}{l} \times \frac{1}{2} + \frac{k-1}{k} \times \frac{l-1}{l} \times \frac{1}{2} \\
&= \Sigma_{x=i,ii,iii,iv} \Pr(\text{Case } x) \times \Pr(Z_{iA}(R) = 1 | \rho_{iA}^{\varphi} = \rho, \theta_i = \theta_{k+1}, \text{Case } x) \\
&= P(Z_{iA}(R) = 1 | \rho_{iA}^{\varphi} = \rho, \theta_i = \theta_{k+1}),
\end{aligned}
$$

where the key inequality is by $l > k$. Therefore, at the above problem, there is no gDA mechanism with any lottery structure for which the qualification IV research design extracts a random assignment. □

B Additional Discussion

B.1 General Definition of a Random Assignment

Definitions 2 and 5 in the main body define a random assignment under the first-choice and qualification instrumental variable (IV) research designs, respectively. This section explains these definitions are special cases of a unified definition of a random assignment under general empirical research designs. It is therefore legitimate to compare the first-choice and qualification IV research designs based on Definitions 2 and 5.

Given any dataset from any assignment problem, I consider a class of empirical research designs that try to identify causal effects of being assigned to any given treatment school s. Each research design in this class tries to achieve the goal by instrumenting for the treatment assignment ($D_{is}(r_0)$) with some instrumental variable $Z_{is}^*(r_0)$, where r_0 is the realized lottery number profile in the data. For simplicity, I consider only binary instrumental variables, i.e., $Z_{is}^*(\cdot) : \mathcal{R} \to \{0, 1\}$ where $\mathcal{R}$ is the set of all possible lottery number profiles. Define Θ as the set of all possible student types $\theta_i = (\succ_i, (\rho_{is})_{s \in S})$. Let $(\Theta_1, ..., \Theta_m)$ be a partition of Θ with $\cup_{k=1}^m \Theta_k = \Theta$. I allow the research design to do instrumenting conditional on which type partition cell contains each student's type (i.e., conditional on $(1\{\theta_i \in \Theta_k\})_{k=1,...,m}$) and within a restricted sample $I_s^*(r_0) \subset I$, where $I_s^*(\cdot) : \mathcal{R} \to \mathcal{I}$ where $\mathcal{I}$ is the set of all subsets of the set of students I.

For outcome Y_i of interest, the research design measures the effect of treatment assignment $D_{is}(r_0)$ on Y_i by estimating the following Two Stage Least Square regression model or a similar IV model within the restricted sample $I_s^*(r_0)$:

$$Y_i = \alpha_2 + \beta_2 D_{is}(r_0) + \Sigma_{k=1}^{m} \gamma_2^k 1\{\theta_i \in \Theta_k\}) + \epsilon_{2i} \text{ (second stage regression)}$$
$$D_{is}(r_0) = \alpha_1 + \beta_1 Z_{is}^*(r_0) + \Sigma_{k=1}^{m} \gamma_1^k 1\{\theta_i \in \Theta_k\}) + \epsilon_{1i} \text{ (first stage regression)}$$

The above class of research designs is parametrized by what IV to use (Z_{is}^*), which aspects of student type to control for $(\Theta_1, ..., \Theta_m)$, and what sample restriction to impose (I_s^*). I allow these objects to change depending on different gDA mechanisms. For any research design in the class, I introduce the following definition of extracting a random assignment.

Definition 8. An empirical research design with instrumental variable Z_{is}^*, conditioning $(\Theta_1, ..., \Theta_m)$, and sample restriction I_s^* **extracts a random assignment** for a gDA mechanism φ if at any assignment problem X and for any school s,

$$P(Z_{is}^*(R) = 1 | i \in I_s^*(r), (1\{\theta_i \in \Theta_k\})_{k=1,...,m} = v, \theta_i = \theta)$$
$$= P(Z_{is}^*(R) = 1 | i \in I_s^*(r), (1\{\theta_i \in \Theta_k\})_{k=1,...,m} = v),$$

for any potential lottery realizations r, any vector $v \in \{0,1\}^m$, and any student type θ for which these conditional probabilities are well-defined.

Only under this conditionally random assignment does the IV Z_{is}^* generate an exogenous or random variation in assignment treatment D_{is} (Heckman and Vytlacil (2007) chapter 4, Manski (2008) chapter 3, Angrist and Pischke (2009) chapter 4). Whether a research design extracts a random assignment depends on which gDA mechanism generates the data since different mechanisms produce different Z_{is}^*, $(\Theta_1, ..., \Theta_m)$, and I_s^*.

The first-choice and qualification IV research designs are two members of this research design class. The first-choice design corresponds to a research design with the treatment assignment as the instrumental variable $Z_{is}^*(r) = D_{is}(r)$, no conditioning $(\Theta_1, ..., \Theta_m) = \Theta$, and sample restriction $I_s^*(r) = First_s(r)$. The qualification IV design corresponds to a research design with the qualification instrumental variable $Z_{is}^*(r) = Z_{is}(r)$, modified priority conditioning $\Theta_k = \{\theta \in \Theta | \rho_{is}^{\varphi} = k\}$, and no sample restriction $I_s^*(r) = I$. Substituting these corresponding objects shows that Definition 8 nests as special cases Definitions 2 and 5 for the first-choice and qualification IV research designs, respectively. Definitions 2 and 5 are therefore comparable, making it legitimate to use these definitions to compare the first-choice and qualification IV research designs.

B.2 Strategy-proofness for Schools is not Exactly Necessary

Section 3.4 shows that the first-choice research design does not extract a random assignment for the DA, Charlotte, and top trading cycles mechanisms, which are not strategy-proof for schools. This suggests strategy-proofness for schools is almost necessary for the first-choice research design to extract a random assignment. On the other hand, strategy-proofness for schools turns out to be not exactly necessary. To see this, consider the following mechanism. Given an assignment problem and realized lottery numbers, the **partially deferred acceptance mechanism** is defined through the following algorithm.[33]

- Step 1: Each student i applies to her most preferred acceptable school (if any). Each school accepts its highest-priority (with respect to $\rho_{is}+r_{is}$) students up to its capacity and rejects every other student. Finalize these acceptances and subtract the number of each school's acceptances from that school's capacity.

- Step 2: Each student who has not been accepted by any school applies to her most preferred acceptable school that has not rejected her (if any). Each school tentatively keeps the highest-ranking students up to its remaining capacity (after the subtraction at step 1), and rejects every other student.

In general, for any step $t \geq 3$,

- Step t: Each student i who was not tentatively assigned to any school in Step $t-1$ applies to her most preferred acceptable school that has not rejected her (if any). Each school tentatively keeps the highest-ranking students up to its remaining capacity (after the subtraction at step 1) from the set of students tentatively assigned to this school in previous step $t-1$ and the students newly applying, and rejects every other student.

The algorithm terminates at the first step at which no student applies to any school. Each student tentatively kept by a school at that step or accepted by that school at step 1 is allocated a seat in that school, resulting in an assignment.

The partially deferred acceptance mechanism is a mix of the Boston mechanism and the DA mechanism in that the first step is process as in the Boston mechanism while the remaining steps are processed as in the DA mechanism. The partially deferred acceptance mechanism can be interpreted as modifying priorities so that each school prioritizes students ranking it first over students ranking it lower, and the partially deferred acceptance mechanism is a gDA mechanism with $(f^{\varphi}(m)=m, g^{\varphi}(n)=1\{n\neq 1\}(K+1))$.

The partially deferred acceptance mechanism is not strategy-proof for schools by a similar reason for the DA mechanism. However, the first-choice research design extracts a random assignment for the partially deferred acceptance mechanism with any lottery regime. The

[33] Agarwal and Somaini (2015) call this mechanism the "first preferences first" mechanism while the same name appears to be used by others to mean a different mechanism. I use "partially deferred acceptance" to avoid confusion.

reason is that given any assignment problem and lottery realization, the treatment assignments of students ranking the treatment school first are finalized at the first step of the algorithm, and their treatment assignments (whether each of them is assigned to the first-choice treatment school) are the same as those produced by the Boston mechanism with the same lottery realization. Corollary 1(a) therefore implies that the first-choice design extracts a random assignment under the partially deferred acceptance mechanism with any lottery regime. Hence, the first-choice research design may extract a random assignment even for a mechanism that is not strategy-proof for schools.

Nevertheless, I am not aware of any empirical study that uses data from the partially deferred acceptance mechanism. As long as more widely-used and widely-discussed mechanisms such as the Boston, DA, Charlotte, and top trading cycles mechanisms are concerned, strategy-proof for schools is necessary, as summarized in Proposition 1.