NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Somouaoga Bonkoungou, Alexander Nesterov

REFORMS MEET FAIRNESS CONCERNS IN SCHOOL AND COLLEGE ADMISSIONS

BASIC RESEARCH PROGRAM
WORKING PAPERS

SERIES: ECONOMICS
WP BRP 234/EC/2020

REFORMS MEET FAIRNESS CONCERNS IN SCHOOL AND COLLEGE ADMISSIONS

SOMOUAOGA BONKOUNGOU AND ALEXANDER NESTEROV

Abstract. Recently, many matching systems around the world have been reformed. These reforms responded to objections that the matching mechanisms in use were unfair and manipulable. Surprisingly, the mechanisms remained unfair even after the reforms: the new mechanisms may induce an outcome with a *blocking student* who desires and deserves a school which she did not receive.

However, as we show in this paper, the reforms introduced matching mechanisms which are *more* fair compared to the counterfactuals. First, most of the reforms introduced mechanisms that are *more fair by stability*: whenever the old mechanism does not have a blocking student, the new mechanism does not have a blocking student either. Second, some reforms introduced mechanisms that are *more fair by counting*: the old mechanism always has at least as many blocking students as the new mechanism.

These findings give a novel rationale to the reforms and complement the recent literature showing that the same reforms have introduced less manipulable matching mechanisms. We further show that the fairness and manipulability of the mechanisms are strongly logically related.

Keywords: market design, school choice, college admission, fairness, stability
JEL Classification: C78, D47, D78, D82

1. Introduction

In the last decades, there has been a wave of reforms of matching systems around the world, ranging from college admissions systems in Chinese provinces, secondary

Higher School of Economics, St.Petersburg
E-mail addresses: bkgsom@gmail.com, nesterovu@gmail.com.
Date: July 23, 2020.
We are grateful to Battal Doğan and Rustamdjan Hakimov for their suggestions. We thank Bettina Klaus, Camille Terrier, Inacio Bo, Madhav Raghavan, and the other participants of the online seminar of the Lausanne Market Design group for their comments, as well as the participants of the online Conference on Mechanism and Institution Design 2020.
Support from the Basic Research Program of the National Research University Higher School of Economics is gratefully acknowledged. The research is partially financed by the RFBR grants 20-01-00687, 19-01-00762.

public school admissions systems in multiple districts in Ghana, to public school admissions systems in multiple cities in the US and in the UK.

The old matching systems were criticized because they were vulnerable to gaming and were unfair. The most vivid example is, perhaps, the 2007 major reform in England, which covers 146 local school admissions systems. According to the Secretary of State, Alan Johnson, the aim of the reform was to "ensure that admission authorities – whether local authorities or schools – operate in a fair way" (School Admissions Code, 2007). The reform prohibited the practice of giving "priority to children according to the order of other schools named as preference by their parents," known as the first-preference-first principle. According to this principle, a student who ranks a school higher in her list, receives a higher admission priority at this school compared to the students who rank this school lower. Prior to the reform, as many as one third of the schools in England used this principle.

In 2009, the Chicago authorities replaced the so-called Boston mechanism that used the same first-preference-first principle for each selective high school, arguing that, due to this principle "high-scoring kids were being rejected simply because of the order in which they listed their college prep preferences" (Pathak and Sönmez, 2013). The same mechanism has been in use for college admission in all provinces in China. It raised similar complaints. For example, one parent said: "My child has been among the best students in his school and school district. He achieved a score of 632 in the college entrance exam last year. Unfortunately, he was not accepted by his first choice. After his first choice rejected him, his second and third choices were already full. My child had no choice but to repeat his senior year" (Chen and Kesten, 2017; Nie, 2007).

The two complaints above illustrate an unfairness issue with the first-preference-first principle. This principle can induce a matching with a so-called *blocking student*, that is, a student who prefers a school over her matching while at least one seat of this school has been assigned to a student with a lower priority (or even left empty). The blocking student desires and deserves this seat, yet she has not been assigned to it. A matching with no blocking student is called *stable*. We use these two concepts to compare the mechanisms before and after each reform.

The first fairness criterion is stability. We compare the mechanisms at each instance, taking the preferences as the reports. A mechanism is **more fair by stability** than a second mechanism if it induces a stable matching whenever the second mechanism induces a stable matching, and the reverse is not true. Namely, for some markets this mechanism induces a stable matching, while the second mechanism does not.

Our main result supports that most of the reforms have adopted matching mechanisms that are more fair by stability. For example, in China, this is true for half of its provinces (Chen and Kesten, 2017). In Chicago, the mechanism adopted after the 2009 reform is more fair by stability than the one previously used (Theorem 1); the one adopted after the 2010 reform is also more fair by stability than the mechanism adopted in 2009 (Theorem 2).

The only exception is the 2007 reform in England in the districts where only some of the schools used the first-preference-first principle. For each of these districts, there are instances where the matching was stable under the old mechanism but yet is not under the new mechanism (Example 1). However, we restored the result when schools in such a district have a common priority order (e.g., based on students' grades or a single lottery), that is, the mechanism adopted after the reform is more fair by stability than the one previously used (Proposition 1).

We formulated a second fairness criterion based on counting the number of blocking students. A mechanism is **more fair by counting** (the number of blocking students) than a second mechanism if for each instance the second mechanism has at least as many blocking students as the first mechanism. This criterion implies the previous one.

Our main result for this criterion supports few reforms showing that the mechanisms adopted after the reforms are more fair by counting than the ones used before. Broadly, these reforms involve extending ranking constraints in the Gale-Shapley mechanism. The Gale-Shapley mechanism with a shorter ranking constraint has more, or an equal number of, blocking students than the Gale-Shapley mechanism with a longer ranking constraint (Theorem 4). This reform took place in Chicago (2010), in Ghana (2007, 2008), in Newcastle (2010), and Surrey (2010) (Pathak and Sönmez, 2013). But the criterion is too demanding for the other reforms. We provide a counterexample showing that after these reforms the number of blocking students may increase (examples 3,4).

Overall, our results provide a new justification for the reforms, complementing the existing ones. Pathak and Sönmez (2013) were the first to observe these reforms and proposed a way to explain them using a notion of *manipulability* that compares mechanisms according to the inclusion of instances where they are not vulnerable to gaming. These results were further strengthened for other mechanisms and other vulnerability criteria (Chen and Kesten, 2017; Decerf and Van der Linden, 2018; Dur et al., 2018; Bonkoungou and Nesterov, 2020).

We also find a logical relationship between stability and manipulability. Under the constrained Gale-Shapley mechanism, when its outcome is stable the mechanism is not manipulable; while for the constrained Boston mechanism, when the mechanism is not manipulable, its outcome is stable (Corollary 1 and Figure 1). For the serial dictatorship mechanism used in Chicago after 2009, the two concepts are equivalent: its outcome at an instance is stable if and only if the mechanism is not manipulable at this instance (Proposition 2).

Another interesting example of the relationship between stability and manipulability is the reform in England. After this reform, the mechanisms in most school districts did not become less manipulable (Bonkoungou and Nesterov, 2020); they also did not become more fair by stability either (Example 1 below). However, for each instance, the reform was successful according to at least one of the two criteria (Proposition 3): if the reform disrupted fairness — by producing an unstable matching while it was stable before the reform — the new matching is not vulnerable to gaming. Thus, at these instances, the mechanisms were not vulnerable after the reform.

The rest of the paper is organized as follows. Next, we briefly review the related literature not mentioned earlier. We then describe the model in section 2. We present the results on the comparisons in section 3 and on the relationship between stability and manipulability in section 4. We present all the proofs in the appendix.

Related literature. Apart from the papers studying the reforms mentioned earlier (Pathak and Sönmez, 2013; Chen and Kesten, 2017; Decerf and Van der Linden, 2018; Bonkoungou and Nesterov, 2020) the recent literature has been interested in other ways to compare mechanism by fairness and stability.

Among the strategy-proof and Pareto efficient mechanisms, the Gale's Top Trading Cycles mechanism (Shapley and Scarf, 1974) is the most fair by stability when each school has one seat (Abdulkadiroglu et al., 2019). This result also holds for other fairness comparisons, such as the set of blocking students (Dogan and Ehlers, 2020b) and the set of blocking triplets (i, j, s) – student i blocking the matching of school s and student j (Kwon and Shorrer, 2019). In fact, the result holds for each stability comparison that satisfies few basic properties (Dogan and Ehlers, 2020b).

Among the Pareto efficient mechanisms, the most fair by stability is the Efficiency-adjusted Deferred Acceptance mechanism (EADA) due to Kesten (2010) – both in terms of blocking pairs and blocking triplets (Dogan and Ehlers, 2020a; Tang and

Zhang, 2020; Kwon and Shorrer, 2019). Independent from the present work, Dogan and Ehlers (2020a) also use the fairness by counting criterion to show that among efficient mechanisms EADA is not the most fair by counting, unless the priority profile satisfies few acyclicity conditions. To our knowledge, this is the only paper that uses the number of blocking students as a measure of fairness.

The first paper that studied the constrained mechanisms is Haeringer and Klijn (2009). They study the stability of the Nash equilibrium outcomes of the game induced by these mechanisms. The most important insight is that the Nash equilibrium outcomes of the constrained Boston mechanism are all stable, while the Nash equilibrium outcomes of constrained Gale-Shapley may not be all stable.[1] In addition, the Nash equilibrium outcomes of a constrained Gale-Shapley are subset of the Nash equilibrium outcomes of any constrained Gale-Shapley with longer list. Therefore, when the Nash equilibrium outcomes the constrained Gale-Shapley with longer list are all stable, the Nash equilibrium outcomes of the constrained Gale-Shapley with shorter list are also stable. Our results do not contradict the spirit of the previous results because in Nash equilibrium students are best responding such that the equilibrium behavior alters the fairness properties of the mechanisms.

2. Model

In a school choice model (Balinski and Sönmez, 1999; Abdulkadiroğlu and Sönmez, 2003), there is a finite and non-empty set I of students with a generic element i and a finite and non-empty set S of schools with a generic element s.

Each student i has a strict preference relation P_i over $S \cup \{\emptyset\}$, where $\emptyset$ represents the outside option for this student. For each student i, let R_i denote the "at least as good as" relation associated with P_i.[2] School s is acceptable to student i if $s \; P_i \; \emptyset$; and it is unacceptable to student i if $\emptyset \; P_i \; s$. The list $P = (P_i)_{i \in I}$ is a preference profile. Given a proper subset $I' \subsetneq I$ of students, we will often write a preference profile as $P = (P_{I'}, P_{-I'})$ to emphasize the components for the students in I'.

Each school s has a strict priority order $\succ_s$ over the set I of students, and a capacity q_s (a natural number indicating the number of its available seats). The list $\succ = (\succ_s)_{s \in S}$ is a priority profile and $q = (q_s)_{s \in S}$ is a capacity vector. We extend each priority order $\succ_s$ of school s to the set 2^I of subsets of students and assume that this

[1] Ergin and Sönmez (2006) also showed that the Nash equilibrium outcomes of the Boston mechanism are stable.

[2] That is, for each $s, s' \in S \cup \{\emptyset\}$, $s \; R_i \; s'$ if and only $s \; P_i \; s'$ or $s = s'$.

extension is responsive to the priority order $\succ_s$ over I as follows. The priority order $\succ_s$ of school s is responsive (Roth, 1986) if

- for each $i, j \in I$ and each $I' \subset I \setminus \{i, j\}$ such that $|I'| < q_s - 1$, we have, (i) $I' \cup \{i\} \succ_s I'$, and (ii) $I' \cup \{i\} \succ_s I' \cup \{j\}$ if and only if $i \succ_s j$ and
- for each $I' \subset I$ such that $|I'| > q_s$, we have $\emptyset \succ_s I'$.

The tuple $(I, S, P, \succ, q)$ is a school choice problem. We assume that there are more students than schools, that is, $|I| > |S|$. The set of students and the set of schools are fixed throughout the paper, and we denote the school choice problem by the triple $(P, \succ, q)$, or even by the preference profile P only.

A matching μ is a function $\mu : I \to S \cup \{\emptyset\}$ such that for each school s, $|\mu^{-1}(s)| \leq q_s$. If $\mu(i) \neq \emptyset$, then we say that student i is matched under μ. If $\mu(i) = \emptyset$, then we say that she is unmatched under μ.

Let $(P, \succ, q)$ be a problem. A matching μ is **individually rational** under P if for each student i, $\mu(i)\ R_i\ \emptyset$. A pair (i, s) of a student and a school blocks the matching μ under $(P, \succ, q)$ if $s\ P_i\ \mu(i)$ and either there is a student j such that $\mu(j) = s$ and $i \succ_s j$ or $|\mu^{-1}(s)| < q_s$. Student i is a **blocking student** for the matching μ under $(P, \succ, q)$ if there is a school s such that the pair (i, s) blocks μ under $(P, \succ, q)$. A matching μ is **stable** under $(P, \succ, q)$ if it is individually rational under P and no pair of a student and a school blocks it.

A **mechanism** φ is a function which maps each school choice problem to a matching. For each problem $(P, \succ, q)$, let $\varphi_i(P, \succ, q)$ denote the component for student i. A mechanism is individually rational if for each problem $(P, \succ, q)$, $\varphi(P, \succ, q)$ is individually rational under P. Given a mechanism φ and a problem $(P, \succ, q)$, we say that $\varphi(P, \succ, q)$ is stable whenever $\varphi(P, \succ, q)$ is stable under $(P, \succ, q)$.

2.1. **Mechanisms.**

We are interested in the mechanisms that were used either before or after the reforms. We first describe the unconstrained versions.

Gale-Shapley. Gale and Shapley (1962) showed that for each problem, there exists a stable matching. In addition, there is a student-optimal stable matching, which is a matching that each student finds at least as good as any other stable matching. For each problem $(P, \succ, q)$, this matching can be found via the Gale and Shapley (1962) student-proposing deferred acceptance algorithm.

- *Step 1:* Each student applies to her most-preferred acceptable school (if any). If a student did not rank any school acceptable, then she remains unmatched. Each school s considers its applicants I_s^1 at this step and tentatively accepts

$\min(q_s, |I_s^1|)$ of the $\succ_s$-highest priority applicants and rejects the remaining ones. Let A_s^1 denote the set of students whom school s has tentatively accepted at this step.

- Step $t (t > 1)$: Each student, who is rejected at step $t-1$, applies to her most-preferred acceptable school among those which have not yet rejected her (if any). If a student does not have any remaining acceptable school, then she remains unmatched. Each school s considers the set $A_s^{t-1} \cup I_s^t$, where I_s^t are its new applicants at this step, and tentatively accepts $\min(q_s, |A_s^{t-1} \cup I_s^t|)$ of the $\succ_s$-highest priority applicants and rejects the remaining ones. Let A_s^t denote the set of students whom school s has tentatively accepted at this step.

The algorithm stops when every student is either accepted at some step or has applied to all of her acceptable schools. The tentative acceptances become final at this step. Let $GS(P, \succ, q)$ denote the matching obtained.

Serial Dictatorship. When schools have the same priority order, we call the Gale-Shapley mechanism the serial dictatorship mechanism.[3] Let $SD(P, \succ, q)$ denote the matching assigned by the serial dictatorship mechanism to the problem $(P, \succ, q)$.

First-Preference-First. The schools are exogenously divided into two subsets $S^{fpf} \subset S$ and $S^{ep} \subset S$ such that they are disjoint and $S^{fpf} \cup S^{ep} = S$. The set S^{eq} is a set of **equal-preference schools** and S^{fpf} is a set of **first-preference-first** schools. The First-Preference-First mechanism (FPF) assigns to each problem $(P, \succ, q)$, the matching $GS(P, \hat{\succ}, q)$, where the priority order of each equal-preference school is maintained intact while the priority order of each first-preference-first school is adjusted according to the rank that students have assigned to this school. Formally, the priority profile $\hat{\succ}$ is obtained as follows:

1. for each equal-preference school $s \in S^{ep}$, $\hat{\succ}_s = \succ_s$ and
2. for each first-preference-first school $s \in S^{fpf}$, $\hat{\succ}_s$ is defined as follows. Let $I^1(s)$ be the set of students who have ranked school s first under P, $I^2(s)$ the set of students who have ranked school s second under P, and so on. Note that we count the ranking of $\emptyset$ as well.

- for each $\ell, k \in \{1, \ldots, |S|+1\}$ such that $\ell > k$ and each students i, j such that $i \in I^k(s)$ and $j \in I^\ell(s)$, $i \hat{\succ}_s j$.
- for each $k \in \{1, \ldots, |S|+1\}$ and each $i, j \in I^k(s)$, $i \hat{\succ}_s j$ if and only if $i \succ_s j$.

[3] According to our definition, a mechanism has as a domain the set of all problems — including problems where schools have different priorities.

Let $FPF(P, \succ, q)$ denote the matching assigned to the problem $(P, \succ, q)$.

Boston. Until 2005, the Boston public school system was using an immediate acceptance mechanism called the Boston mechanism (Abdulkadiroğlu and Sönmez, 2003). This mechanism assigns to each problem $(P, \succ, q)$, the matching as described in the following algorithm.

- *Step 1:* Each student applies to her most-preferred acceptable school (if any). Each school s, considers its applicants I_s^1 at this step and immediately accepts $\min(q_s, |I_s^1|)$ of the $\succ_s$-highest priority applicants and rejects the remaining ones. For each school s, let $q_s^1 = q_s - \min(q_s, |I_s^1|)$ denote its remaining capacity after this step.
- *Step t:* $(t > 1)$ Each student who is rejected at step $t-1$, applies to her most-preferred acceptable school among those which have not yet rejected her (if any). Each school s considers its new applicants I_s^t at this step and immediately accepts $\min(q_s^{t-1}, |I_s^t|)$ of the $\succ_s$-highest priority applicants and rejects the remaining ones. For each school s, let $q_s^t = q^{t-1} - \min(q_s^{t-1}, |I_s^t|)$ denote its remaining capacity after this step.

The algorithm stops when every student is either accepted at some step or has applied to all of her acceptable schools. Let $\beta(P, \succ, q)$ denote the matching assigned by the Boston mechanism to the problem $(P, \succ, q)$.

Remark. *In the (algorithm of the) Boston mechanism, students applying to the same school at each step have assigned the same rank to it. Therefore, students applying to a school at a given step of the algorithm rank this school higher than those applying to it at any step after. In particular, no student could be rejected by a school while another student, who has assigned a lower rank to it, is accepted. Thus, the Boston mechanism is a first-preference-first mechanism where every school is a first-preference-first school. This result follows from the Proposition 2 of Pathak and Sönmez (2008).*

Constrained mechanisms. Haeringer and Klijn (2009) first observed that in practice, students are allowed to report a limited number of schools. This means that schools that are listed below a certain position are not considered. Let $k \in \{1, \ldots, |S|\}$. For each student i, the truncation after the k'th acceptable school (if any) of the preference relation P_i with x acceptable schools is the preference relation P_i^k with $\min(x, k)$ acceptable schools such that all schools are ordered as in P_i. Let $P^k = (P_i^k)_{i \in I}$. The constrained version φ^k of the mechanism φ is the mechanism that assigns to each problem $(P, \succ, q)$ the matching $\varphi(P^k, \succ, q)$. That is, $\varphi^k(P, \succ, q) = \varphi(P^k, \succ, q)$.

Chinese parallel. Chen and Kesten (2017) describe a parametric mechanism that many Chinese provinces have been using. The parameter $e \geq 1$ is a natural number. For each problem $(P, \succ, q)$, the outcome is a sequential application of constrained GS. In the first round, the matching is final for matched students under $GS^e(P, \succ, q)$, while unmatched students proceed to the next round. In the next round, each school reduces its capacity by the number of students assigned to it in the last round, each matched student replaces her preferences with a preference relation where she finds no school acceptable and the unmatched students (in the previous round) are matched according to GS^{2e} for the reduced capacities and the new preference profile. The process continues until either no school has a remaining seat or no unmatched student finds a school with a remaining seat acceptable. Let $Ch^{(e)}(P, \succ, q)$ denote the matching assigned by the mechanism to $(P, \succ, q)$.[4]

3. Results

3.1. **Fairness by stability.** Our starting point is a comparison according to the set inclusion of the problems where mechanisms are stable.

Definition 1 (Chen and Kesten, 2017). *Mechanism φ' is* ***more fair by stability*** *than φ if*

(i) at each problem where φ is stable, φ' is also stable and

(ii) there exists a problem where φ' is stable but φ is not.

Although this criterion is less demanding, in the sense that it does not take into account the problems where mechanisms produce unstable outcomes, it does not explain all reforms. Indeed, it does not explain many changes that followed the 2007 reform in the UK. Indeed, the constrained First-Preference-First mechanism is not comparable to the constrained Gale-Shapley mechanism according to this concept. We demonstrate this in the following example.

Example 1. *Let $I = \{i_1, \ldots, i_7\}$ and $S = \{s_1, \ldots, s_5\}$. Let school s_3 be a (the only) first-preference-first school. Let $(P, \succ, q)$ be a problem where each school has one seat and the remaining components are specified as follows. The sign $\vdots$ indicates that the remaining part is arbitrary.*

[4]This definition of the Chinese parallel mechanisms is given only for the symmetric version where each round has the same length e. See Chen and Kesten (2017) for details.

P_{i_1}	P_{i_2}	P_{i_3}	P_{i_4}	P_{i_5}	P_{i_6}	P_{i_7}	$\succ_{s_1}$	$\succ_{s_2}$	$\succ_{s_3}$	$\succ_{s_4}$	$\succ_{s_5}$
s_1	s_1	s_4	s_1	s_2	s_1	s_5	i_4	i_5	i_3	i_1	i_7
s_2	s_3	s_3	s_2	s_1	s_2	s_1	$\vdots$	$\vdots$	i_1	i_6	$\vdots$
s_3	$\emptyset$	$\emptyset$	s_3	s_3	s_5	s_2			i_2	i_3	
s_4			$\emptyset$	$\emptyset$	s_3	$\emptyset$			$\vdots$	$\vdots$	
$\emptyset$					s_4						
					$\emptyset$						

The outcomes of the constrained First-Preference-First FPF^4 and the constrained Gale-Shapley GS^4 at $(P, \succ, q)$ are specified as follows:

$$FPF^4(P, \succ, q) = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 & i_6 & i_7 \\ s_4 & \emptyset & s_3 & s_1 & s_2 & \emptyset & s_5 \end{pmatrix},$$

$$GS^4(P, \succ, q) = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 & i_6 & i_7 \\ s_3 & \emptyset & s_4 & s_1 & s_2 & \emptyset & s_5 \end{pmatrix}.$$

The matching $FPF^4(P, \succ, q)$ is stable.[5] *However, the matching $GS^4(P, \succ, q)$ is not stable. Indeed, the pair (i_6, s_4) blocks this matching because student i_6 is unmatched and finds school s_4 acceptable, but student i_3 is matched to s_4 while $i_6 \succ_{s_4} i_3$.*

The intuition is that the constraint in GS shortened the chains of the rejections needed to reach a stable matching in the Gale-Shapley algorithm. For example, student i_3 is temporarily matched to school s_4 at some step of the algorithm. At the student-optimal stable matching for $(P, \succ, q)$, school s_4 is assigned to student i_1. However, we need an application of student i_1 at that school to displace student i_3 from s_4. This does not occur under GS^4 because no student initiates the rejection chain. However, under FPF^4, the application of student i_2 at school s_3 causes the rejection of student i_1 at s_3 (student i_2 has ranked it higher than i_1 and school s_3 is a first-preference-first school). This is the rejection needed to reach the student-optimal stable matching.

In this example, we illustrate how the constrained GS mechanism has shortened the chains needed to reach a stable matching. It is well known that this type of chains cause welfare losses of the unconstrained GS (Kesten, 2010).[6] However, under the Boston mechanism, (where all schools are first-preference-first schools) there is no such chain. The following result is an implication of this fact.

[5]Note that this matching is the student (and school)-optimal stable matching.

[6]These chains are initiated by the so-called interrupters. These are students who initiate chains of rejections which return to them (Kesten, 2010).

Theorem 1. *Suppose that there are at least two schools and let $k > 1$. The constrained Gale-Shapley mechanism GS^k is more fair by stability than the constrained Boston mechanism β^k.*

See the appendix for the proof. Similarly, when schools have a common priority order, there is no such chain in the Gale-Shapley mechanism. We restore the result for this case.

Proposition 1. *Suppose that there are at least two schools and at least one first-preference-first school, and let $k > 1$. The constrained serial dictatorship mechanism SD^k is more fair by stability than the constrained First-Preference-First mechanism FPF^k.*

See the appendix for the proof. The constrained GS with shorter and longer lists mechanisms can also be compared with this criterion. However, the intuition for this result is different. When the constrained Gale-Shapley with shorter lists is stable, the restriction has no effect on the outcome.

Lemma 1. *Let $(P, \succ, q)$ be a problem and $k > 1$. Then $GS^k(P, \succ, q)$ is stable if and only if $GS^k(P, \succ, q) = GS(P, \succ, q)$.*

See the appendix for the proof. Then, when the constraint in GS^k does not affect the outcome, the longer constraint in GS^{k+1} will not affect the outcome either.

Theorem 2. *Suppose that there are at least three schools and let $k > \ell$ where k is less than the number of schools and $\ell \geq 1$. Then, the constrained Gale-Shapley mechanism GS^k is more fair by stability than GS^ℓ.*

See the appendix for the proof. Finally, we consider the Chinese mechanisms. These mechanisms are known to be comparable in terms of fairness by stability, but only in case one tier is a multiple of another (Chen and Kesten, 2017). We present this result for completeness.

Theorem 3 (Chen and Kesten, 2017)**.** *For each $e \geq 1, m > 1$, the Chinese mechanism $Ch^{(me)}$ is more fair by stability than $Ch^{(e)}$.*

3.2. **Fairness by counting.** In this section we present the results for a stronger comparison criterion. Unlike the previous notion, we compare the number of blocking students. Therefore, the mechanisms are compared for all problems (even where they induce unstable outcomes).

Definition 2. *An individually rational mechanism φ' is* ***more fair by counting*** *(the blocking students) than an individually rational mechanism φ if*

(i) *for each problem, there are at least as many blocking students of the outcome of φ as there are of the outcome of φ', and*
(ii) *there is a problem where there are more blocking students of the outcome of φ than the outcome of φ'.*

Fairness by counting is a stronger notion than stability considered earlier. If a mechanism φ' is more fair by counting than φ, then for each problem where φ induces a stable matching, i.e., there is no blocking student, φ' also necessarily induces a stable matching. Our main result with this concept is a strengthening of the comparison between different constraints of the Gale-Shapley mechanism.

We illustrate the intuition using the example below.

Example 2. *Let $I = \{i_1, \ldots, i_5\}$ and $S = \{s_1, \ldots, s_4\}$. Let $(P, \succ, q)$ be a problem where each school has one seat, and the remaining components are specified as follows.*

P_{i_1}	P_{i_2}	P_{i_3}	P_{i_4}	P_{i_5}	$\succ_{s_1}$	$\succ_{s_2}$	$\succ_{s_3}$	$\succ_{s_4}$
s_1	s_1	s_2	s_3	s_3	i_3	i_2	i_1	i_5
s_2	s_2	s_1	s_1	s_4	i_1	i_4	i_5	$\vdots$
s_3	s_3	s_3	s_2	$\vdots$	$\vdots$	$\vdots$	$\vdots$	

Let us compare the mechanisms GS^2 and GS^1. We have

$$GS^2(P, \succ, q) = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 \\ \emptyset & s_2 & s_1 & \emptyset & s_3 \end{pmatrix}$$

where student i_1 is the unique blocking student for the matching under $(P, \succ, q)$. Indeed, she is unmatched, finds s_3 acceptable while she has a higher priority at s_3 than i_5. Let us shorten the reported list only for student i_2. Then,

$$GS^2(P^1_{i_2}, P_{-i_2}, \succ, q) = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 \\ s_1 & \emptyset & s_2 & \emptyset & s_3 \end{pmatrix}.$$

As a result of this replacement, there are three types of students, given their status in the previous matching. First, student i_2 — who was matched — became a blocking student. Second, student i_1 — who was a blocking student — is not a blocking student for the new matching. Finally, student i_4 is a new blocking student.

The intuition of this result is that by shortening the schools listed by student i_2, she is worse off while the other students are weakly better off. First, she is a blocking student for the new matching. Second, student i_1 is not a blocking student for the new matching, though she was a blocking student for the old matching. But a new blocking student appears so that there are two blocking students in total.

This turns out to be true in general. When a student shortens the list, the set of blocking students changes, but the size of this set never decreases (and sometimes increases). And when all students shorten their lists, we get the following result.

Theorem 4. *Suppose that there are at least two schools and let $k > \ell$ where k is less than the number of schools and $\ell \geq 1$. The constrained Gale-Shapley mechanism GS^k is more fair by counting than GS^ℓ.*

See the appendix for the proof. Next, we show that the other comparisons do not extend to this stronger criterion. The first example shows that the constrained Boston mechanism is not comparable to the constrained GS.

Example 3 (Constrained Boston and GS)**.** *Let $n \geq 7$, $I = \{i_1, ..., i_n\}$ and $S = \{s_1, \ldots, s_5\}$. Let $(P, \succ, q)$ be a problem where each school has one seat and the remaining components are specified as follows.*

P_{i_1}	P_{i_2}	P_{i_3}	P_{i_4}	P_{i_5}	$\ldots$	$P_{i_{n-1}}$	P_{i_n}	$\succ_{s,\, s\in S}$
s_1	s_2	s_3	s_1	s_1	s_1	s_1	s_4	i_1
$\vdots$	$\vdots$	$\vdots$	s_4	s_2	s_2	s_2	s_5	i_2
			s_5	s_3	s_3	s_3	$\vdots$	i_3
			$\vdots$	s_5	s_5	s_5		i_4
				$\emptyset$	$\emptyset$	$\emptyset$		i_5
								$\vdots$
								i_n

The outcomes of β^3 and GS^3 for this problem are specified as follows:

$$\beta^3(P, \succ, q) = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 & \ldots & i_{n-1} & i_n \\ s_1 & s_2 & s_3 & s_5 & \emptyset & \ldots & \emptyset & s_4 \end{pmatrix}$$

and

$$GS^3(P, \succ, q) = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 & i_5 & \ldots & i_{n-1} & i_n \\ s_1 & s_2 & s_3 & s_4 & \emptyset & \ldots & \emptyset & s_5 \end{pmatrix}.$$

Let us compare the number of blocking students for the two outcomes. On one hand, student i_4 is the only blocking student for $\beta^3(P, \succ, q)$. Indeed, the pair (i_4, s_4)

Reforms	From	To	more fair by stability?		more fair by counting?	
			Arbitrary priority	Common priority	Arbitrary priority	Common priority
UK(54), 2007/11	FPF^k	GS^k	not comparable	more	not comparable	not comparable
Chicago, 2009 UK(4), 2007	β^k	GS^k	more	more	not comparable	not comparable
Chicago, 2010 Ghana, 2007/08 UK(2), 2010	GS^k	GS^{k+1}	more	more	more	more
China(13), 2001/12	β	$Ch^{(e)}$	more	more	not comparable	not comparable

Table 1. Comparison of the matching mechanisms by fairness criteria.

Notes: Each row compares the mechanism in the third column to the mechanism in the second column with respect to fairness by stability and fairness by counting. Common priority is a special case of arbitrary priority. The complete list of the UK local matching systems and Chinese provinces that underwent the reforms can be found in Pathak and Sönmez (2013) and, respectively, in Chen and Kesten (2017).

blocks $\beta^3(P, \succ, q)$ under $(P, \succ, q)$. On the other hand, students $i_5, \ldots, i_{n-1}$ are all blocking students for $GS^3(P, \succ, q)$ because they are unmatched, each of them prefers school s_5 to being unmatched, and has higher priority than i_n under $\succ_{s_5}$. Since $n \geq 7$, there are at least two blocking students for $GS^3(P, \succ, q)$. Therefore, there are more blocking students for $GS^3(P, \succ, q)$ than $\beta^3(P, \succ, q)$. Under Theorem 1, there is a problem where GS^3 is stable but not β^3.

Next, the symmetric Chinese parallel mechanisms are also not comparable in terms of fairness by counting.

Example 4 (Chinese parallel)**.** *We consider Example 3. Consider the Chinese mechanisms $Ch^{(1)} = \beta$ and $Ch^{(3)}$ and note that for the problem $(P, \succ, q)$ specified in that example, $Ch^{(1)}(P, \succ, q) = \beta^3(P, \succ, q)$ and $Ch^{(3)}(P, \succ, q) = GS^3(P, \succ, q)$. According to the conclusion in Example 3, there are more blocking students for $Ch^{(3)}(P, \succ, q)$ than $Ch^{(1)}(P, \succ, q)$. According to Chen and Kesten (2017), there is a problem where $Ch^{(3)}$ produces a stable outcome but $Ch^{(1)}$ does not.*

The overall ranking with respect to the two criteria are presented in Table 1. In the next section, we investigate the relationship between stability and manipulability.

Remark. *Dogan and Ehlers (2020a) introduced a criterion where they compare mechanisms by the inclusion of the blocking pairs and blocking students. However, these criteria are stronger than fairness by counting and will lead to negative results for our comparisons. To see this, consider Example 3. In this example,* (i_5, s_5) *is a blocking pair of* $SD^4(P, \succ, q)$ *but not* $\beta^4(P, \succ, q)$*. In addition,* (i_4, s_4) *is a blocking pair of* $\beta^4(P, \succ, q)$ *but not* $SD^4(P, \succ, q)$*.*

For the comparison between different constrained Gale-Shapley, consider Example 2. There, (i_1, s_3) *is a blocking pair of* $GS^2(P, \succ, q)$ *but not* $GS^1(P, \succ, q)$*. In addition,* (i_2, s_2) *is a blocking pair of* $GS^1(P, \succ, q)$ *but not* $GS^2(P, \succ, q)$*.*

4. Stability and manipulability

In this section, we will elucidate the relation between blocking students and manipulating students, i.e., those who may benefit from misrepresenting their preferences to the mechanisms. We define those students below as well as the definition of a non-manipulable mechanism.

Definition 3. *Let* φ *be a mechanism.*

(i) Student i *is a* ***manipulating student*** *of* φ *at* $(P, \succ, q)$ *if there is a preference relation* $\hat{P}_i$ *such that*

$$\varphi_i(\hat{P}_i, P_{-i}, \succ, q) \; P_i \; \varphi_i(P, \succ, q).$$

(ii) The mechanism φ *is* ***not manipulable*** *at* $(P, \succ, q)$ *if there is no manipulating student of* φ *at* $(P, \succ, q)$*.*

It turns out that there is a strong relation between blocking students and manipulating students for the constrained Boston mechanism and the constrained Gale-Shapley mechanism. Interestingly, these relations for the two mechanisms are reversed.

Theorem 5. *Let* $(P, \succ, q)$ *be a problem and* $k > 1$*. Then,*

(i) every blocking student of the outcome $\beta^k(P, \succ, q)$ *of the constrained Boston mechanism is a manipulating student of* β^k *at* $(P, \succ, q)$ *and*

(ii) every manipulating student of the constrained Gale-Shapley mechanism GS^k *at* $(P, \succ, q)$ *is a blocking student of* $GS^k(P, \succ, q)$*.*

See the appendix for the proof. These results have important implications for the relation between manipulability and stability. To see this, suppose that there is no manipulating student for the constrained Boston mechanism β^k at $(P, \succ, q)$. Then, under part (i) of this theorem, there is not a blocking student of $\beta^k(P, \succ, q)$. Since

β^k is individually rational, then $\beta^k(P, \succ, q)$ is stable. Suppose now that there is no blocking student for $GS^k(P, \succ, q)$. Since GS^k is individually rational, this means that $GS^k(P, \succ, q)$ is stable. Then, there is no manipulating student for GS^k at $(P, \succ, q)$. That is, GS^k is not manipulable at $(P, \succ, q)$. We summarize these results in the following corollary; see the appendix for the proof.

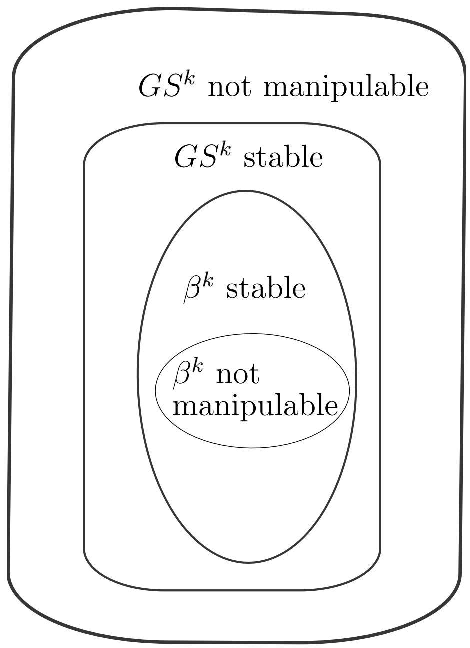

Figure 1. The set inclusion relations of the problems where GS^k and β^k are stable or not manipulable.

Corollary 1. *Let $(P, \succ, q)$ be a problem and $k > 1$. (i) Suppose that the constrained Boston mechanism β^k is not manipulable at $(P, \succ, q)$. Then, $\beta^k(P, \succ, q)$ is stable.*

(ii) Suppose that the constrained Gale-Shapley mechanism $GS^k(P, \succ, q)$ is stable. Then, GS^k is not manipulable at $(P, \succ, q)$.

Note that there are problems where the reverse of each of these results does not hold. See Example 5 below for the constrained Gale-Shapley mechanism. To see a counterexample of the reverse of the case (i), consider a problem $(P, \succ, q)$ where students have a common ranking of schools, have ranked k schools acceptable and where each school has one seat. Then, $\beta^k(P, \succ, q)$ is stable. However, the student who has received her third ranked school is better off top ranking the school she has ranked second as her top choice.

The diagrams illustrate the structure of the interplay between stability and manipulability for the constrained Boston and the constrained Gale-Shapley mechanisms.

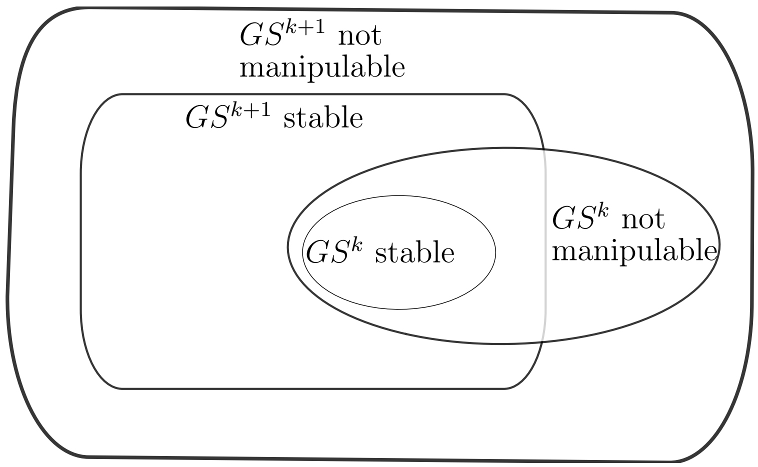

Figure 2. The set inclusion relations of the problems where GS^k and GS^{k+1} are stable or not manipulable.

An implication of the latter results is a manipulability comparison introduced by Pathak and Sönmez (2013). Under part (i) of Corollary 1, when the constrained Boston mechanism is not manipulable then it is stable. By Theorem 1, the constrained Gale-Shapley mechanism is also stable. By part (ii) of Corollary 1, the constrained Gale-Shapley mechanism is not manipulable. This is the comparison established by Pathak and Sönmez (2013).

Corollary 2. *(Pathak and Sönmez, 2013). Let $(P, \succ, q)$ be a problem, $k > 1$ and suppose that the constrained Boston mechanism β^k is not manipulable at $(P, \succ, q)$. Then, the constrained Gale-Shapley mechanism GS^k is not manipulable at $(P, \succ, q)$.*

Another implication is for the serial dictatorship mechanism. The manipulation strategy under the constrained GS is to include an acceptable school in the list. But when the constrained GS is stable, all the seats of such a school are assigned to higher priority students, and such a manipulation does not help. This implies that constrained serial dictatorship mechanism is non-manipulable and stable for the same set of problems.

Proposition 2. *Let $(P, \succ, q)$ be a problem and $k > 1$. The constrained serial dictatorship mechanism SD^k is stable if and only if it is not manipulable at $(P, \succ, q)$.*

See the appendix for the proof. In general, the constrained Gale-Shapley mechanism may be unstable while not manipulable. We illustrate this in the following example.

Example 5. *Let $I = \{i_1, \ldots, i_4\}$ and $S = \{s_1, \ldots, s_4\}$. Let $(P, \succ, q)$ be a problem where each school has one seat and the remaining components are specified as follows.*

P_{i_1}	P_{i_2}	P_{i_3}	P_{i_4}	$\succ_{s_1}$	$\succ_{s_2}$	$\succ_{s_3}$	$\succ_{s_4}$
s_1	s_1	s_2	s_3	i_1	i_4	i_3	$\vdots$
$\vdots$	s_2	s_3	s_2	$\vdots$	i_3	i_2	
	s_3	$\vdots$	$\vdots$		i_2	i_4	
	$\emptyset$				i_1	i_1	

Let us consider the constrained Gale-Shapley mechanism GS^2. We have

$$GS^2(P, \succ, q) = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 \\ s_1 & \emptyset & s_2 & s_3 \end{pmatrix}.$$

This matching is not stable under $(P, \succ, q)$ because student i_2 is unmatched, finds school s_3 acceptable while student i_4 is matched to it and $i_2 \succ_{s_2} i_4$. We claim that GS^2 is not manipulable at $(P, \succ, q)$. Only student i_2 could benefit from misrepresenting her preferences to the mechanism GS^2 because each of the other students is matched to her most-preferred school. Let $P_{i_2}^{s_3}$ be a preference relation where student i_2 has ranked only school s_3 acceptable. Then,

$$GS^2(P_{i_2}^{s_3}, P_{-i_2}, \succ, q) = \begin{pmatrix} i_1 & i_2 & i_3 & i_4 \\ s_1 & \emptyset & s_3 & s_2 \end{pmatrix},$$

that is, student i_2 remains unmatched even by ranking school s_3 first. (It is easy to verify that any other manipulation also leaves i_2 unmatched.) Therefore, GS^2 is not manipulable at $(P, \succ, q)$. The intuition is that this ranking initiates a chain of rejections which returns to this student. Student i_2 becomes a so-called "interrupter" when she ranks school s_3 first (Kesten, 2010).

We also establish another direct corollary of Theorem 5 with two additional results. We show that when switching from constrained Boston to constrained GS, of from the constrained GS to a longer list, the mechanism becomes more fair by stability and less manipulable.

Corollary 3. *Let $(P, \succ, q)$ be a problem.*

(i) Let $k > 1$ and suppose that the constrained Boston mechanism β^k is stable at $(P, \succ, q)$. Then, the constrained Gale-Shapley mechanism GS^k is stable and not manipulable at $(P, \succ, q)$.

(ii) Let $k > \ell > 1$ and suppose that the constrained Gale-Shapley mechanism GS^{ℓ} is stable at $(P, \succ, q)$. Then, the mechanism GS^k is stable and not manipulable at $(P, \succ, q)$.

Finally, we partially restore the comparisons for the First-Preference-First mechanism. Although the constrained First-Preference-First mechanism and the constrained Gale-Shapley mechanism are not comparable by manipulability (Bonkoungou and Nesterov, 2020) and by fairness by stability (Example 1), there is a surprising interplay between the two concepts.

Proposition 3. *Let $(P, \succ, q)$ be a problem, $k > 1$ and suppose that the constrained First-Preference-First mechanism FPF^k is stable under $(P, \succ, q)$. Then, the constrained Gale-Shapley mechanism GS^k is not manipulable at $(P, \succ, q)$.*

See the appendix for the proof.

This result helps to evaluate the reforms in England, where FPF^k was replaced by GS^k. Even though for some problems the reform was unsuccessful in one of the two dimensions — decreasing fairness by stability (Example 1) or increase manipulability (Pathak and Sönmez, 2013; Bonkoungou and Nesterov, 2020) — the reform could not be unsuccessful in both dimensions.

To sum up the results of this section, stability and manipulability are logically related, and the relationship depends on the mechanism.

5. Conclusions

In response to objections, many school districts around the world have recently reformed there admissions systems. The main reason for these objections was that the mechanisms were unfair and manipulable. Yet, the mechanisms remained unfair and manipulable even after the reforms. We used two criteria and showed that many reforms resulted in more fair matching mechanisms, first by relying on stability and second by counting and comparing the number of the blocking students.

The reforms concern essentially two major changes. First, they kept the constraint on the number of schools that each student is allowed to report but replaced the Boston mechanism (or a hybrid between Gale-Shapley and Boston mechanism) with the Gale-Shapley's student-proposing deferred acceptance mechanism. Second, some school districts extended the number of schools that each student is allowed to report but kept the Gale-Shapley mechanism. Our findings support these reforms, as well as such changes in the future.

References

Abdulkadiroglu, A., Che, Y.-K., Pathak, P., Roth, A., and Tercieux, O. (2019). Efficiency, justified envy, and incentives in priority-based matching. *American Economic Review: Insights, forthcoming.*

Abdulkadiroğlu, A. and Sönmez, T. (2003). School choice: A mechanism design approach. *American Economic Review*, 93(3):729–747.

Balinski, M. and Sönmez, T. (1999). A tale of two mechanisms: Student placement. *Journal of Economic Theory*, 84(1):73–94.

Bonkoungou, S. and Nesterov, A. S. (2020). Comparing school choice and college admission mechanisms by their strategic accessibility. *Working paper, Higher School of Economics.*

Chen, Y. and Kesten, O. (2017). Chinese college admissions and school choice reforms: A theoretical analysis. *Journal of Political Economy*, 125(1):99–139.

Decerf, B. and Van der Linden, M. (2018). Manipulability in constrained school choice. *Available at SSRN 2809566.*

Department for Education and Skills (2007). School admissions code. *Norwich, United Kingdom: Her Majesty's Stationery Office.*

Dogan, B. and Ehlers, L. (2020a). Blocking pairs versus blocking students: Stability comparisons in school choice. *Available at SSRN.*

Dogan, B. and Ehlers, L. (2020b). Robust minimal instability of the top trading cycles mechanism. *Available at SSRN.*

Dubins, L. E. and Freedman, D. A. (1981). Machiavelli and the Gale-Shapley algorithm. *American Mathematical Monthly*, 88(7):485–494.

Dur, U. M., Pathak, P. A., Song, F., and Sönmez, T. (2018). Deduction dilemmas: The Taiwan assignment mechanism. Technical report, National Bureau of Economic Research.

Ergin, H. and Sönmez, T. (2006). Games of school choice under the Boston mechanism. *Journal of Public Economics*, 90(1):215–237.

Gale, D. and Shapley, L. S. (1962). College admissions and the stability of marriage. *American Mathematical Monthly*, 69(1):9–15.

Gale, D. and Sotomayor, M. (1985). Some remarks on the stable matching problem. *Discrete Applied Mathematics*, 11(3):223–232.

Haeringer, G. and Klijn, F. (2009). Constrained school choice. *Journal of Economic Theory*, 144(5):1921–1947.

Kesten, O. (2010). School choice with consent. *Quarterly Journal of Economics*, 125(3):1297–1348.

Kwon, H. and Shorrer, R. I. (2019). Justified-envy minimal mechanisms in school choice. *Available at SSRN 3495266*.

Nie, H. (2007). A game-theoretical analysis of china's college admission mechanism. *China Economic Quarterly Beijing*, 6(3):899.

Pathak, P. and Sönmez, T. (2008). Levelling the playing field: Strategic and sophisticated players in the Boston mechanism. *American Economic Review*, 98(1):99–123.

Pathak, P. A. and Sönmez, T. (2013). School admissions reform in Chicago and England: Comparing mechanisms by their vulnerability to manipulation. *American Economic Review*, 103(1):80–106.

Roth, A. E. (1982). The economics of matching: Stability and incentives. *Mathematics of Operations Research*, 7(4):617–628.

Roth, A. E. (1986). On the allocation of residents to rural hospitals: A general property of two-sided matching markets. *Econometrica: Journal of the Econometric Society*, pages 425–427.

Shapley, L. and Scarf, H. (1974). On cores and indivisibility. *Journal of Mathematical Economics*, 1(1):23–37.

Tang, Q. and Zhang, Y. (2020). Weak stability and pareto efficiency in school choice. *Economic Theory*, pages 1–20.

Appendix: Proofs

The following result from the literature will be useful throughout the appendix. This result is known as the *rural hospital theorem.*

Lemma 2 (Rural hospital theorem)**.** *(Roth, 1986). Let $(P, \succ, q)$ be a problem, ν and μ two stable matchings. Then,*

(i) the same set of students are matched under ν and μ and

(ii) each school is matched to the same number of students under ν and μ.

Proposition 3: *Let $(P, \succ, q)$ be a problem, $k > 1$ and suppose that the constrained First-Preference-First mechanism FPF^k is stable at $(P, \succ, q)$. Then, the constrained Gale-Shapley mechanism GS^k is not manipulable $(P, \succ, q)$.*

Proof. We first establish two claims.

<u>Claim 1:</u> *Suppose that student* i *is matched to school* s *under* $GS^k(P, \succ, q)$ *and let* P_i^s *be a preference relation where she has ranked only school* s *as an acceptable choice. Then, she is matched to school* s *under* $GS^k(P_i^s, P_{-i}, \succ, q)$.

Suppose that $GS_i^k(P, \succ, q) = s$. As shown by Roth (1982), $GS_i(P^k, \succ, q) = s$ implies that $GS_i(P_i^s, P_{-i}^k, \succ, q) = s$. Since $k > 1$, the truncation of P_i^s after the k'th acceptable school is nothing but P_i^s. Thus, $GS_i^k(P_i^s, P_{-i}, \succ, q) = s$.

<u>Claim 2:</u> *Suppose that student* i *can manipulate* GS^k *at the problem* $(P, \succ, q)$. *Then she is unmatched under* $GS^k(P, \succ, q)$.

This result follows from Pathak and Sönmez (2013).

We are now ready to prove the proposition. Let $(P, \succ, q)$ be a problem and suppose that $\mu = FPF^k(P, \succ, q)$ is stable under $(P, \succ, q)$. Under Claim 2, it is enough to show that an unmatched student under $GS^k(P, \succ, q)$ has no profitable misrepresentation. Because GS^k is individually rational, under Claim 1, we need to restrict ourselves to manipulation by top ranking a school first. Since μ is stable under $(P, \succ, q)$, then μ is also stable under $(P^k, \succ, q)$. Since GS^k is individually rational, we need to check that there is no blocking pair. Suppose, to the contrary, that a pair (i, s) is a blocking pair for μ under $(P^k, \succ, q)$. Then, $s \ P_i^k \ \mu(i)$ and either (i) school s has an empty seat under μ or (ii) there is a student j such that $\mu(j) = s$ and $i \succ_s j$. Note that $s \ P_i^k \ \mu(i)$ implies that $s \ P_i \ \mu(i)$. Therefore, (i, s) is also a blocking pair for μ under $(P, \succ, q)$. This conclusion contradicts our assumption that μ is stable under $(P, \succ, q)$.

Therefore, μ is stable under $(P^k, \succ, q)$. Since $GS(P^k, \succ, q)$ is the student-optimal stable matching under $(P^k, \succ, q)$,

$$\text{for each student } i, \ GS_i(P^k, \succ, q) \ R_i^k \ \mu(i). \tag{1}$$

In line with Lemma 2, the same number of students are matched under μ and $GS(P^k, \succ, q)$. Let i be a student and s a school and suppose that i is unmatched under $GS(P^k, \succ, q)$ and that $s \ P_i \ GS_i(P^k, \succ, q)$. Then, student i is also unmatched under μ. Thus, $s \ P_i \ \mu(i) = \emptyset$. Because μ is stable under $(P, \succ, q)$, this implies that every student in $\mu^{-1}(s)$ has higher priority than i under $\succ_s$. Let P_i^s denote a preference relation where i has ranked only school s acceptable. Since μ is stable under $(P^k, \succ, q)$, it is also stable under $(P_i^s, P_{-i}^k, \succ, q)$. Under Lemma 2, the set of matched students is the same at all stable matchings. Thus, student i is also unmatched under $GS(P_i^s, P_{-i}^k, \succ, q)$. Then, under Claim 1, there is no strategy P_i' such that $GS_i^k(P_i', P_{-i}) = s$. Thus, the mechanism GS^k is not manipulable at $(P, \succ, q)$. $\square$

<u>Theorem 1:</u> *Suppose that there are at least three schools and let $k > 1$. The constrained Gale-Shapley mechanism GS^k is more fair by stability than the constrained Boston mechanism β^k.*

Proof. The Boston mechanism is a special case of the First-Preference-First mechanism when every school is a first-preference-first school. Let $(P, \succ, q)$ be a problem and suppose that $\beta^k(P, \succ, q)$ is stable under $(P, \succ, q)$. As stated in equation 1, each student finds the outcome $GS^k(P, \succ, q)$ at least as good as $\beta^k(P, \succ, q)$ under P^k. We also know that the Boston mechanism is Pareto efficient, that is, for each problem there is no other matching that each student finds at least as good as its outcome (Abdulkadiroğlu and Sönmez, 2003). Therefore, the matching $\beta^k(P, \succ, q) = \beta(P^k, \succ, q)$ is Pareto efficient under P^k. Thus, $GS^k(P, \succ, q) = \beta^k(P, \succ, q)$ and consequently, $GS^k(P, \succ, q)$ is stable under $(P, \succ, q)$.

We construct a problem where GS^k is stable but not β^k. Since there are at least two schools and more students than schools, let s_1, s_2 be two distinct schools and i_1, i_2 and i_3 three students. Let $(P, \succ, q)$ be a problem where each school has one seat and the remaining components are specified as follows.

$P_{i \neq 3}$	P_3	$\succ_{s \in S}$
s_1	s_2	i_1
s_2	s_1	i_2
$\emptyset$	$\emptyset$	i_3
		$\vdots$

Since $k \geq 2$, $GS^k(P, \succ, q) = GS(P, \succ, q)$ is stable under $(P, \succ, q)$. However, the matching

$$\beta^k(P, \succ, q) = \begin{pmatrix} i_1 & i_3 & i \neq 1, 3 \\ s_1 & s_2 & \emptyset \end{pmatrix}$$

is not stable because the pair (i_2, s_2) blocks it under $(P, \succ, q)$.

□

<u>Proposition 1:</u> *Suppose that there are at least two schools and at least one first-preference-first school, and let $k > 1$. The constrained serial dictatorship mechanism SD^k is more fair by stability than the constrained First-Preference-First mechanism FPF^k.*

Proof. Let $(P, \succ, q)$ be a problem where schools have a common priority order and suppose that $FPF^k(P, \succ, q)$ is stable under $(P, \succ, q)$. Under equation 1, each student finds the outcome $SD^k(P, \succ, q)$ at least as good as $FPF^k(P, \succ, q)$ under P^k. With a common priority order, there is a unique stable matching under $(P, \succ, q)$ which is also Pareto efficient under P. Therefore, because $FPF^k(P, \succ, q)$ is stable under $(P, \succ, q)$, we have $FPF^k(P, \succ, q) = SD(P, \succ, q)$.

Next, every student who is matched under $SD(P, \succ, q)$ is matched to one of her top k-ranked acceptable schools. Therefore, $SD(P, \succ, q) = FPF^k(P, \succ, q)$ is also Pareto efficient under P^k. Thus, equation 1 implies that $SD^k(P, \succ, q) = FPF^k(P, \succ, q)$ and consequently, $SD^k(P, \succ, q)$ is stable under $(P, \succ, q)$.

We can adapt the example provided in the proof of Theorem 1 to show that there is a problem where SD^k is stable but not FPF^k. □

Lemma 1: *Let $(P, \succ, q)$ be a problem and $k > 1$. Then $GS^k(P, \succ, q)$ is stable if and only if $GS^k(P, \succ, q) = GS(P, \succ, q)$.*

Proof. The "if" part is straightforward because $GS(P, \succ, q) = GS^k(P, \succ, q)$ is the student-optimal stable matching under $(P, \succ, q)$.

The "only if" part. Suppose that $GS^k(P, \succ, q)$ is stable under $(P, \succ, q)$. Let $N = \{i \in I | GS_i(P, \succ, q) = \emptyset\}$ denote the set of students who are unmatched under $GS(P, \succ, q)$.

Step 1: For each $i \in N$, $GS_i(P, \succ, q) = GS_i(P^k, \succ, q)$.

This follows from the assumption that $GS(P^k, \succ, q)$ is stable and Lemma 2.

Step 2: For each $i \in I \setminus N$, $GS_i(P, \succ, q) = GS_i(P^k, \succ, q)$.

Let $i \in I \setminus N$. Because $GS(P, \succ, q)$ is the student-optimal stable matching under $(P, \succ, q)$,

$$GS_i(P, \succ, q) \; R_i \; GS_i^k(P, \succ, q). \tag{2}$$

Note that for each student $j \in N$, the preference relation P_j can be interpreted as if she has extended her list of acceptable schools from P_j^k. As shown by Gale and Sotomayor (1985), when a subset of students extend their list of acceptable schools, none of the remaining students are better off. Therefore,

$$\text{for each student } j \in I \setminus N, \;\; GS_j(P_N^k, P_{-N}, \succ, q) \; R_j \; GS_j(P, \succ, q). \tag{3}$$

Because GS is individually rational under P, under equation 3, student i is also matched under $GS(P^k_N, P_{-N}, \succ, q)$. Next, since $GS(P^k, \succ, q)$ is stable under $(P, \succ, q)$, by assumption, Lemma 2 implies that the same set of students are matched under both $GS(P, \succ, q)$ and $GS(P^k, \succ, q)$. Therefore, i is also matched under $GS(P^k, \succ, q)$. Next, note that the students in $I \setminus N$ have extended their list of acceptable schools under (P^k_N, P_{-N}) from P^k. Then, at the end of the Gale-Shapley algorithm for the problem $(P^k, \succ, q)$, each of the students in $I \setminus N$ is accepted by a school. The school that each of them has listed below the school that has accepted her at this step of the algorithm and how she has ranked them do not affect the outcome of the algorithm. Thus,

$$GS(P^k, \succ, q) = GS(P^k_N, P_{-N}, \succ, q).$$

This equation and equation 3 imply that $GS_i(P^k, \succ, q)\ R_i\ GS_i(P, \succ, q)$. Since the preference relation P_i is strict, this relation and equation 2 imply that

$$GS_i(P^k, \succ, q) = GS_i(P, \succ, q).$$

Finally, by *Step 1* and *Step 2*, the matching is the same for each student under $GS^k(P, \succ, q)$ and $GS(P, \succ, q)$, the desired conclusion. □

<u>*Corollary 1:*</u> *Let $(P, \succ, q)$ be a problem and $k > 1$.*

(i) *Suppose that the constrained Boston mechanism β^k is not manipulable at the problem $(P, \succ, q)$. Then, $\beta^k(P, \succ, q)$ is stable.*
(ii) *Suppose that the constrained Gale-Shapley mechanism GS^k is stable under $(P, \succ, q)$. Then, GS^k is not manipulable at $(P, \succ, q)$.*

Proof. We prove (i) by the contraposition. Suppose that $\beta^k(P, \succ, q)$ is not stable under $(P, \succ, q)$. Since β^k is individually rational, there is a pair (i, s) of a student and a school which blocks $\beta^k(P, \succ, q)$ under $(P, \succ, q)$. In assent with (i) of Theorem 5, student i is a manipulating student of β^k at $(P, \succ, q)$. Thus, β^k is manipulable at $(P, \succ, q)$.

We now prove part (ii). Suppose that $GS^k(P, \succ, q)$ is stable under $(P, \succ, q)$. As shown in Lemma 1, we have $GS^k(P, \succ, q) = GS(P, \succ, q)$. As shown by Pathak and Sönmez (2013), only unmatched students could benefit from misrepresenting their preferences to GS^k. Let i be an unmatched student under $\mu = GS^k(P, \succ, q)$, s a school such that $s\ P_i\ \mu(i)$ and P_i^s a preference relation where she has ranked only school s as an acceptable school. Because μ is stable under $(P, \succ, q)$, each student in $\mu^{-1}(s)$ has a higher priority than i under $\succ_s$. Therefore, μ is also stable under $(P_i^s, P_{-i}^k, \succ, q)$. Under Lemma 2, student i is unmatched under both μ and $GS(P_i^s, P_{-i}^k, \succ, q)$. Corresponding to Claim 1, there is no strategy P_i' such that student i is matched to school s under $GS^k(P_i', P_{-i}, \succ, q)$. Therefore, student i cannot manipulate GS^k at $(P, \succ, q)$. Since the choice of i is arbitrary, GS^k is not manipulable at $(P, \succ, q)$. $\square$

<u>Theorem 4</u>: *Let $k > \ell$ and suppose that there are at least k schools. The constrained Gale-Shapley mechanism GS^k is more fair by counting than GS^ℓ.*

Proof. Let $(P, \succ, q)$ be a problem. The strategy is to start from $GS(P^k, \succ, q)$ and replace the preference relations in P^k one at a time, with a preference relation in P^ℓ for the corresponding student until we get $GS(P^\ell, \succ, q)$. We prove the theorem by showing that the number of blocking students is not decreasing after each replacement.

First, we prove two lemmas. The first lemma will be used at some steps of the second. The second will be the main part to proving the theorem.

Lemma 3. *Let N be a subset of students and $\mu = GS(P_N^\ell, P_{-N}^k, \succ, q)$. Any blocking student for μ under $(P, \succ, q)$ is unmatched.*

Proof. We prove the lemma by the contradiction. Suppose, to the contrary, that some student i is a blocking student for $\mu = GS(P_N^\ell, P_{-N}^k, \succ, q)$ under $(P, \succ, q)$ such that $\mu(i) = s$ for some school s. Then, there is a school s' such that $s'\ P_i\ \mu(i)$ and either (i) $|\mu^{-1}(s')| < q_{s'}$ or (ii) there is a student j such that $\mu(j) = s'$ and $i \succ_{s'} j$. Let $x \in \{\ell, k\}$ be such that $x = \ell$ if $i \in N$ and $x = k$ if $i \notin N$. Since $\mu(i) = s$, school s is one of the top x acceptable schools under P_i. Thus $s'\ P_i^x\ \mu(i) = s$. This relation, together with the case (i) or (ii) imply that the pair (i, s') blocks the matching μ under $(P_N^\ell, P_{-N}^k, \succ, q)$. This conclusion contradicts the fact that μ is stable under $(P_N^\ell, P_{-N}^k, \succ, q)$.

$\square$

The next lemma is the main part for proving the theorem. For simplicity, let $\hat{P} = (P^{\ell}_N, P^k_{-N})$, $i \notin N$ and denote

$$\mu =: GS(\hat{P}, \succ, q)$$

and,

$$\nu = GS(P^{\ell}_i, \hat{P}_{-i}, \succ, q).$$

Lemma 4. *There are at least as many blocking students for ν than μ under $(P, \succ, q)$.*

Proof. Let n be the number of blocking students for μ under $(P, \succ, q)$. We show that there are at least n blocking students for ν under $(P, \succ, q)$.

Let us first show that every student other than i finds ν at least as good as μ under $\hat{P}$. To see this, note that student i has extended her list of acceptable schools under $\hat{P}_i = P^k_i$ from P^{ℓ}_i. As shown by Gale and Sotomayor (1985), after this extension no student, other than i, is better off. That is,

$$\text{for each student } j \neq i,\ \ \nu(j)\ \hat{R}_j\ \mu(j). \tag{4}$$

We divide the rest of the proof into two cases. In the first case, student i is unmatched under μ. For this case, we will show that any blocking student for μ under $(P, \succ, q)$ is also a blocking student for ν under $(P, \succ, q)$. In the second case, student i is matched under μ. We will show that either $\mu = \nu$ (in case i is also matched under ν), or i is a blocking student for ν.

Case I: Suppose that student i is unmatched under μ, that is, $\mu(i) = \emptyset$.

We first show that $\nu(i) = \emptyset$. Suppose, to the contrary, that $\nu(i) = s$, for some school s. Then s is one of the top ℓ acceptable schools of student i under P_i. Since $k > \ell$, school s is also one of the top k acceptable schools under $\hat{P}_i = P^k_i$. Therefore,

$$GS_i(P^{\ell}_i, \hat{P}_{-i}, \succ, q) = s\ \hat{P}_i\ \mu(i) = GS_i(\hat{P}, \succ, q) = \emptyset.$$

This relation shows that student i is better off misrepresenting her preference to the Gale-Shapley mechanism, contradicting the fact that this mechanism is not manipulable (Dubins and Freedman, 1981; Roth, 1982). Therefore, $\nu(i) = \mu(i) = \emptyset$. This equality together with equation 4 imply that each student finds the matching ν at least as good as μ under $\hat{P}$. Because $\mu = GS(\hat{P}, \succ, q)$ is stable under $(\hat{P}, \succ, q)$, it is also stable under $(P_i^\ell, \hat{P}_{-i}, \succ, q)$. To see this, note that student i is unmatched under μ and that for each school s such that $s\ P_i^\ell\ \mu(i)$, school s does not have an empty seat under μ and every student in $\mu^{-1}(s)$ has higher priority than i under $\succ_s$. Since $\nu = GS(P_{-i}^\ell, \hat{P}_{-i}, \succ, q)$ is also stable under $(P_{-i}^\ell, \hat{P}_{-i}, \succ, q)$, under Lemma 2, we have the following conclusion.

Conclusion: (a) the same set of students are matched (unmatched) under ν and μ and (b) every school is matched to the same number of students under both ν and μ.

Let us now prove that every blocking student for μ under $(P, \succ, q)$ is also a blocking student for ν under $(P, \succ, q)$. Let j be a blocking student for μ under $(P, \succ, q)$. There are two cases.

Case I.1: $j = i$

Then, there is a school s such that $s\ P_i\ \mu(i)$ and either (i) school s has an empty seat under μ or (ii) there is a student j' such that $\mu(j') = s$ and $i \succ_s j'$.

Consider the case (i) where school s has an empty seat under μ. Then, under part (b) of the previous conclusion, s has an empty seat under ν. Since $\nu(i) = \emptyset$, i is a blocking student for ν under $(P, \succ, q)$.

Consider the case (ii) where there is a student j' such that $\mu(j') = s$ and $i \succ_s j'$. Without loss of generality, suppose that school s does not have an empty seat under μ. Then, under part (b) of the previous conclusion, school s does not have an empty seat under ν. Suppose that $\nu(j') = s$. Since $\nu(i) = \emptyset$ and $i \succ_s j'$, then the pair (i, s) blocks ν under $(P, \succ, q)$ and i is a blocking student for ν under $(P, \succ, q)$. Suppose that $\nu(j') \neq s$. Since $|\nu^{-1}(s)| = q_s$, there is $j'' \in \nu^{-1}(s) \setminus \mu^{-1}(s)$. Under equation 4,

$$s = \nu(j'')\ \hat{P}_{j''}\ \mu(j'').$$

Since $\mu = GS(\hat{P}, \succ, q)$ is stable under $(\hat{P}, \succ, q)$ and $\mu(j') = s$, then $j' \succ_s j''$. Because $\succ_s$ is transitive, $i \succ_s j'$ and $j' \succ_s j''$ imply that $i \succ_s j''$. Since $s\ P_i\ \nu(i) = \emptyset$, the pair (i, s) blocks ν under $(P, \succ, q)$ and i is a blocking student for ν under $(P, \succ, q)$.

Case I.2: $j \neq i$

There is a school s such that $s\ P_j\ \mu(j)$ and either (i) school s has an empty seat under μ or (ii) there is a student j' such that $\mu(j') = s$ and $j \succ_s j'$. As shown in Lemma 3, because student j is a blocking student for μ under $(P, \succ, q)$, we have $\mu(j) = \emptyset$.

Let us consider the case (i). Under Lemma 2, student j is also unmatched under ν and school s has an empty seat under ν. Thus, j is a blocking student for ν under $(P, \succ, q)$.

Second, consider (ii) where there is a student j' such that $\mu(j') = s$ and $j \succ_s j'$. If $\nu(j') = s$, then (j, s) is a blocking pair of ν under $(P, \succ, q)$ and j is a blocking student for ν under $(P, \succ, q)$. Without loss of generality, suppose that $|\mu^{-1}(s)|= q_s$ and $\nu(j') \neq s$. According to part (b) of the above conclusion, $|\nu^{-1}(s)|= q_s$. Then, there is a student $j'' \in \nu^{-1}(s) \setminus \mu^{-1}(s)$. Because $\mu(i) = \emptyset$, we have $j'' \neq i$, and by equation 4,

$$s = \nu(j'')\ \hat{P}_{j''}\ \mu(j'').$$

Since μ is stable under $(\hat{P}, \succ, q)$ and $\mu(j') = s$, then this equation implies that $j' \succ_s j''$. Because $\succ_s$ is transitive, $j \succ_s j'$ and $j' \succ_s j''$ imply that $j \succ_s j''$. Since $s\ P_j\ \nu(j) = \emptyset$, then the pair (j, s) blocks ν under $(P, \succ, q)$ and j is a blocking student for ν under $(P, \succ, q)$.

In conclusion, every blocking student for μ under $(P, \succ, q)$ is also a blocking student for ν under $(P, \succ, q)$. There are n blocking students for ν under $(P, \succ, q)$.

Case II: Student i is matched under μ.

If student i is also matched under $\nu = DA(P_i^\ell, \hat{P}_{-i}, \succ, q)$, then $\nu = \mu$. To see this, let $\nu(i) = s$ for some school s. School s is one of the top ℓ acceptable schools of student i under P_i. The Gale-Shapley mechanism is invariant to the modification of the preferences of the students for the part below their outcomes. We know that P_i^k is one such modification of P^ℓ below school s. Thus, $\nu = \mu$. We now consider the case where i is not matched under ν.

Suppose that student i is unmatched under ν. The strategy of the proof is to show that i is a blocking student for ν under $(P, \succ, q)$ and that there are also at least $n-1$ other blocking students for ν under $(P, \succ, q)$. We depict the flow of these students in Figure 3.

Step 1: Student i is a blocking student for ν under $(P, \succ, q)$.

Recall that we assumed that student i is matched under $\mu = GS(\hat{P}, \succ, q)$, where $\hat{P} = (P^{\ell}_{N}, P^{k}_{-N})$ and $i \notin N$. Let $s = \mu(i)$. School s is one of the top k acceptable schools under P_i. Since $\nu(i) = \emptyset$, if school s has an empty seat under ν, then clearly the pair (i, s) blocks ν under $(P, \succ, q)$ and i is a blocking student for ν under $(P, \succ, q)$. Suppose that $|\nu^{-1}(s)| = q_s$. Since $\mu(i) = s$ and $\nu(i) = \emptyset$, there is a student $j \in \nu^{-1}(s) \setminus \mu^{-1}(s)$. By equation 4,

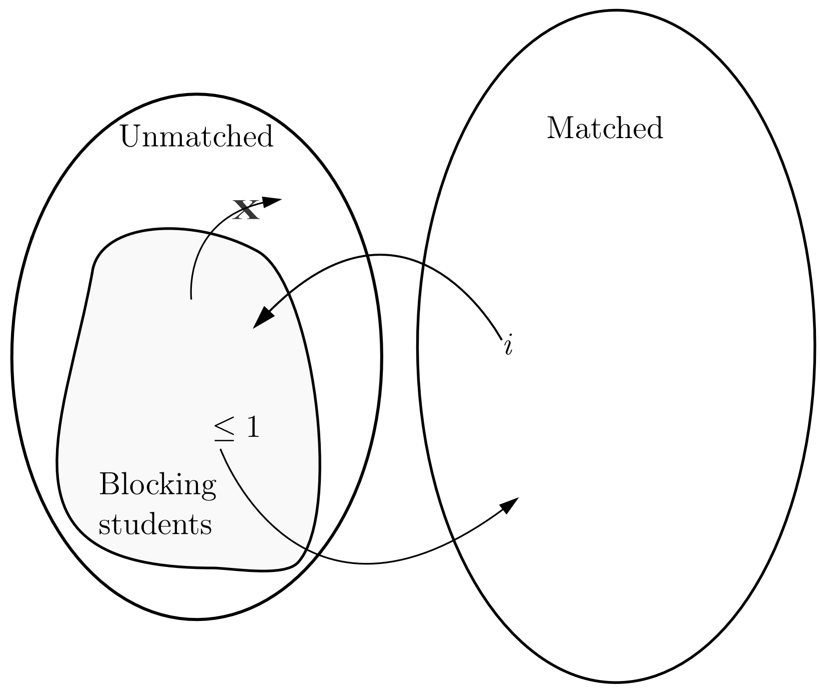

Figure 3. Case II: flow of students across matched, unmatched, and blocking status, from μ to ν: at most one student can leave the blocking status to the matched status; student i left the matched status to the blocking status, and no student can leave the blocking status and remain unmatched.

$$s = \nu(j) \; \hat{P}_j \; \mu(j).$$

Since μ is stable under $(\hat{P}, \succ, q)$ and $\mu(i) = s$, we have $i \succ_s j$. Therefore, the pair (i, s) blocks ν under $(P, \succ, q)$ and i is a blocking student for ν under $(P, \succ, q)$.

Step 2: Every blocking student for μ under $(P, \succ, q)$ who is unmatched under ν is also a blocking student for ν under $(P, \succ, q)$.

Let j be a blocking student for μ under $(P, \succ, q)$ and suppose that she is unmatched under ν. There is a school s such that $s \; P_j \; \mu(j)$ and either (i) school s has an empty seat under μ or (ii) there is a student j' such that $\mu(j') = s$ and $j \succ_s j'$. In addition, because j is the blocking student of μ under $(P, \succ, q)$, by Lemma 3, we have $\mu(j) = \emptyset$.

Let us consider the case (i) where school s has an empty seat under μ. We also show that s has an empty seat under ν. Assume otherwise. Then, there is $j' \in \nu^{-1}(s) \setminus \mu^{-1}(s)$. We know that student i is unmatched under ν. Thus, $j' \neq i$. Under equation 4, $s = \nu(j') \; \hat{P}_{j'} \; \mu(j')$. This contradicts the fact that μ is stable under $(\hat{P}, \succ, q)$ because s has an empty seat under μ. Therefore, s has an empty seat under ν. Then the pair (j, s) blocks ν under $(P, \succ, q)$ and j is a blocking student for ν under $(P, \succ, q)$.

Let us now consider the case (ii) where there is a student j' such that $\mu(j') = s$ and $j \succ_s j'$. If school s has an empty seat under ν, then because student j is unmatched under ν, she is a blocking student for ν under $(P, \succ, q)$. Suppose that school s does not have an empty seat under ν. If $\nu(j') = s$ then the pair (j, s) blocks ν under $(P, \succ, q)$ and j is a blocking student for ν under $(P, \succ, q)$ because $\nu(j) = \emptyset$ and $j \succ_s j'$. Suppose that $\nu(j') \neq s$. Because school s does not have an empty seat under ν, there is $j'' \in \nu^{-1}(s) \setminus \mu^{-1}(s)$. Since student i is unmatched under ν, we have $j'' \neq i$. By equation 4, we have $s = \nu(j'') \; \hat{P}_{j''} \; \mu(j'')$. Since μ is stable under $(\hat{P}, \succ, q)$, the equation and the fact that $\mu(j') = s$ imply that $j' \succ_s j''$. Because $\succ_s$ is transitive, $j \succ_s j'$ and $j' \succ_s j''$ imply that $j \succ_s j''$. Since $s \; P_j \; \nu(j) = \emptyset$, then the pair (j, s) blocks ν under $(P, \succ, q)$ and j is a blocking student for ν under $(P, \succ, q)$.

Step 3: Every student but i who is matched under μ is also matched under ν.

Suppose that for some student $j \neq i$ and some school s, $\mu(j) = s$. Under equation 4, we have $\nu(j) \; \hat{R}_j \; \mu(j) = s$. Since μ is individually rational under $\hat{P}$, $\nu(j) \neq \emptyset$.

Step 4: There are at least n blocking students for ν under $(P, \succ, q)$.

Let j be a blocking student for μ under $(P, \succ, q)$ who is not a blocking student for ν under $(P, \succ, q)$. Then, j is matched under ν. Otherwise, according to step 2, she is also a blocking student for ν under $(P, \succ, q)$. We prove, more generally, that there are at most one student who is unmatched under μ but matched under ν. To do that, we compare for each school the number of students matched to it under μ and ν.

Let s be a school. Suppose that it does not have an empty seat under μ. Then, we have $|\nu^{-1}(s)| \leq |\mu^{-1}(s)| = q_s$. Suppose now that s has an empty seat under μ. Suppose that there is $j' \in \nu^{-1}(s) \setminus \mu^{-1}(s)$. Then, because student i is unmatched under ν, $j' \neq i$. By equation 4,

$$s = \nu(j') \; \hat{P}_{j'} \; \mu(j').$$

This contradicts the fact that μ is stable under $(\hat{P}, \succ, q)$ because school s has an empty seat under μ. Thus, there is no student matched to school s under ν but not under μ. Therefore, $|\nu^{-1}(s)| \leq |\mu^{-1}(s)|$. We conclude that no school is matched to more students under ν than under μ. Thus,

$$\sum_{s \in S} |\nu^{-1}(s)| \leq \sum_{s \in S} |\mu^{-1}(s)|. \tag{5}$$

By step 3, all students, but student i, who are matched under μ are also matched under ν. Therefore, the set of students who are matched under ν consists of the following students:

- the students who are matched under μ, except student i and
- the students who are unmatched under μ but matched under ν.

Let x denote the number of the students who are unmatched under μ but matched under ν. Then, we have

$$\sum_{s \in S} |\nu^{-1}(s)| = \underbrace{\sum_{s \in S} |\mu^{-1}(s)| - 1}_{\substack{\text{number of students} \\ \text{matched under } \mu \text{ and } \nu}} + x,$$

where the first two expressions on the right-hand side indicate that we subtracted student i from those who are matched under μ. By rearranging this equation, we get

$$\sum_{s \in S} |\nu^{-1}(s)| - \sum_{s \in S} |\mu^{-1}(s)| = x - 1 \leq 0,$$

where the inequality follows from equation 5. Thus, there is at most one student who is unmatched under μ but matched under ν. According to Lemma 3, all blocking students for μ under $(P, \succ, q)$ are unmatched under μ. Then, there is at most one blocking student for μ under $(P, \succ, q)$ who is matched under ν. In assent with this result together with step 2, there is at most one blocking student for μ under $(P, \succ, q)$ who is not a blocking student for ν under $(P, \succ, q)$. Among the n blocking students for μ under $(P, \succ, q)$, at most one of them is not a blocking student for ν under

$(P, \succ, q)$. Therefore, excluding student i, there are at least $n-1$ blocking students for ν under $(P, \succ, q)$. Since student i is also a blocking student for ν under $(P, \succ, q)$, there are at least n blocking students for ν under $(P, \succ, q)$.

□

We complete the proof of the theorem by applying the lemma sequentially. Let n be the number of the blocking students for $GS(P^k, \succ, q)$ under $(P, \succ, q)$. For simplicity, let $I = \{1, 2, \ldots, |I|\}$. Under Lemma 4, there are at least n blocking students for

$$\mu_1 = GS(P_1^\ell, P_{-1}^k, \succ, q)$$

under $(P, \succ, q)$. By the same lemma, compared to μ_1, there are at least n blocking students of the matching

$$\mu_2 = GS(P_{\{1,2\}}^\ell, P_{-\{1,2\}}^k, \succ, q)$$

under $(P, \succ, q)$. With a repeated replacement of the remaining components of P^k with their counterparts in P^ℓ, we draw the conclusion that there are at least n blocking students for $GS(P^\ell, \succ, q)$ under $(P, \succ, q)$.

Finally, we describe a problem where the outcome of GS^ℓ has more blocking students than the outcome of GS^k. Let $(P, \succ, q)$ be a problem where each school has one seat, each student has k acceptable schools and such that students have a common ranking of schools. Then, $GS^k(P, \succ, q) = GS(P, \succ, q)$. Thus $GS^k(P, \succ, q)$ is stable under $(P, \succ, q)$. Let s be the school that students have ranked at the k'th position starting from the top. Since there are more students than schools and $k > \ell$, at least one student is not matched under $GS^\ell(P, \succ, q)$ and no student is matched to school s even though every student prefers it to being unmatched. Then, there is at least one blocking student for $GS^\ell(P, \succ, q)$. Therefore, there are more blocking students for $GS^\ell(P, \succ, q)$ than $GS^k(P, \succ, q)$ under $(P, \succ, q)$. □

Proposition 2: *Let $(P, \succ, q)$ be a problem and $k > 1$. The constrained serial dictatorship mechanism SD^k is stable if and only if it is not manipulable at $(P, \succ, q)$.*

Proof. As shown by Bonkoungou and Nesterov (2020), SD^k is not manipulable at $(P, \succ, q)$ if and only if $SD^k(P, \succ, q) = SD(P, \succ, q)$. Suppose that $SD^k(P, \succ, q)$ is stable. Then, according to Lemma 1, $SD^k(P, \succ, q) = SD(P, \succ, q)$ and thus SD^k

is not manipulable at $(P, \succ, q)$. Suppose that SD^k is not manipulable at $(P, \succ, q)$. Then, $SD^k(P, \succ, q) = SD(P, \succ, q)$ and thus stable. $\square$

<u>Theorem 5</u>: *Let $(P, \succ, q)$ be a problem and $k > 1$.*

(i) *Every blocking student of $\beta^k(P, \succ, q)$ is a manipulating student of the constrained Boston mechanism β^k at $(P, \succ, q)$.*
(ii) *Every manipulating student of the constrained Gale-Shapley mechanism GS^k at $(P, \succ, q)$ is a blocking student of $GS^k(P, \succ, q)$.*

Proof. Proof of (i). Let i be a student and suppose that she is a blocking student of $\mu = \beta(P^k, \succ, q)$. There is a school s such that the pair (i, s) blocks μ under $(P, \succ, q)$. Then, we have $s\ P_i\ \mu(i)$ and either (a) school s has an empty seat under μ or (b) there is a student j such that $\mu(j) = s$ and $i \succ_s j$. We claim that student i did not rank school s first under P_i. Otherwise, school s has rejected student i at the first step of the Boston algorithm under $(P^k, \succ, q)$. This is because $k > 1$ and the top ranked schools are considered under β^k. This contradicts the assumption that school s has an empty seat or has accepted student j and $i \succ_s j$. Let P_i^s be a preference relation where i has ranked school s first. Since s has an empty seat under $\beta^k(P, \succ, q)$ or has accepted student j and $i \succ_s j$, there are less than q_s students who have ranked school s first under P^k and have a higher priority than i under $\succ_s$. Therefore, $\beta_i(P_i^s, P_{-i}^k, \succ, q) = s$. Since $s\ P_i\ \mu(i)$, then i is a manipulating student of β^k at $(P, \succ, q)$.

Proof of (ii). We prove this part by contradiction. Suppose that student i is a manipulating student of GS^k at $(P, \succ, q)$ but is not a blocking student for $\mu = GS^k(P, \succ, q)$ under $(P, \succ, q)$. By Claim 2 above, i is unmatched under $GS^k(P, \succ, q)$. Let s be a school such that $s\ P_i\ \mu(i)$. Then, $|\mu^{-1}(s)| = q_s$ and every student in $\mu^{-1}(s)$ has higher priority than i under $\succ_s$. Let P_i^s be a preference relation where i has ranked only school s as an acceptable school. Since μ is stable under $(P^k, \succ, q)$, it is also stable under $(P_i^s, P_{-i}^k, \succ, q)$. This follows from the fact that $\mu(i) = \emptyset$ and that every student in $\mu^{-1}(s)$ has a higher priority than i under $\succ_s$. According to Lemma 2, the set of unmatched students is the same under μ and $GS^k(P_i^s, P_{-i}^k, \succ, q)$. Thus, i is also unmatched under $GS^k(P_i^s, P_{-i}^k, \succ, q)$. According to Claim 1, there is no misreport by which i is matched to s. Since s has been chosen arbitrarily, i is not a manipulating student of GS^k at $(P, \succ, q)$. This conclusion contradicts our assumption that student i is a manipulating student of GS^k at $(P, \succ, q)$. $\square$

<u>Theorem 2</u>: *Suppose that there are at least three schools and let $k > \ell$ where k is less than the number of schools and $\ell \geq 1$. Then, the constrained Gale-Shapley mechanism GS^k is more fair by stability than GS^ℓ.*

Proof. Suppose that $GS^\ell(P, \succ, q)$ is stable under $(P, \succ, q)$. Then, there is no blocking student for it under $(P, \succ, q)$. According to Theorem 4, there is no blocking student for $GS^k(P, \succ, q)$ under $(P, \succ, q)$. Since $GS^k(P, \succ, q)$ is individually rational under P, then it is stable under $(P, \succ, q)$.

We described an example in the proof of Theorem 4 where there is a blocking student (pair) for GS^ℓ but not GS^k. Since GS^k and GS^ℓ are individually rational, at this problem GS^k is stable but not GS^ℓ. $\square$