Ex-ante Fairness in the Boston Mechanism under Pre-exam Preference Submission

Jaimie Lien, Jie Zheng and Xiaohan Zhong

Department of Economics
School of Economics and Management
Tsinghua University

This Version: June 2\textsuperscript{nd}, 2014
Initial Version: January 20\textsuperscript{th}, 2014

Abstract: 1

In an exam-based school choice matching mechanism (e.g., China’s college admission), student exam scores are used to determine school priorities. However, a student’s true qualifications or ability may not be perfectly revealed by his/her exam score. The matching outcome can thus be undesirable in the sense that it is not ex-ante fair, i.e., it may not match students with higher intrinsic abilities to better colleges. A Boston mechanism with the requirement that students submit their preferences over colleges before the exam score is realized (i.e., pre-BOS mechanism) can potentially improve ex-ante fairness. We compare the timing of students’ preference submissions being before versus after the exam under either the Boston (BOS) or Serial dictatorship (SD) mechanisms, under the assumption of homogeneous student preferences. We characterize the equilibrium under which the pre-BOS mechanism can achieve ex-ante fairness. We find that a further restriction requiring that students can only include one school in their submission list (i.e., constrained pre-BOS) can implement the ex-ante fair matching outcome more easily than the unconstrained pre-BOS.

Keywords: Preference Submission Timing, Boston, Serial Dictatorship, Ex-ante Fairness, Constrained School Choice

JEL classification: C78, C92, D81, I28

\footnote{1 jaimie.lien.tsinghuasem@gmail.com; zhengjie@sem.tsinghua.edu.cn; zhongxh@sem.tsinghua.edu.cn; We thank Nanyang Bu, Kang Rong, Ning Sun, Qianfeng Tang, Zaifu Yang, Yongchao Zhang for helpful comments. This research is funded by National Natural Science Foundation of China, Project No.71173127, 71203112, and 71303127. All errors are our own.}
1. Introduction

In a school to student matching problem, the matching result is called fair (alternatively, stable, or free of justified envy) when there is no school-student pair which prefer each other than its matched counterpart. This concept, originating in Gale and Shapley (1962), has become indispensable for evaluating the welfare properties of various matching mechanisms. However, the concept does not presume any particular school priorities, or rules regarding schools’ priorities. Unlike students, who have individually rational preferences over possible schools, school priorities are often determined by social norms or by central planners.

It is often assumed schools prefer “better” students, just as students prefer “better” colleges. But what is a “better” student for a school and why should it prefer a “better” student? A better student is arguably a student who has higher intrinsic abilities, and therefore given the quality and quantity of education provided by the college, he or she would have higher after-education productivity. Schools may prefer more able students because they may establish a better reputation for schools, although this reputation does not come directly from high quality or quantity of education. Such a school priority rule can also earn support from a social welfare perspective. That is, it allows better students to ‘choose’ schools first. Imagine that a Serial dictatorship (SD) mechanism is used to implement a fair matching outcome; the schools' priority problem then becomes: what kind of students should have the right to choose first? Arguably, students with higher abilities should have this priority, particularly if these abilities are acquired through past effort. (This is similar to the notion that a person with higher income would have a larger choice set of consumption goods. Assuming that his income is fairly earned, this rule of distributing goods is generally not challenged by economists.)

In practice, however, students’ abilities are unobservable. A matching mechanism attempting to achieve the aforementioned ability-based fairness criterion, must find observable substitutes or proxies for students’ abilities. An exam is one of the most frequently used proxies: intuitively, students with higher academic exam scores are those students with higher academic abilities. If exam scores are indeed a perfect proxy for abilities, a SD mechanism can be relied on to reach ability-based fair matching outcomes. The problem is that exam scores are often an imperfect measure of students’ actual abilities. An ability-based fairness result may not be achieved under an exam or score-based SD mechanism.

In this paper we study a new mechanism which may be able to achieve improvements in ability-based fairness, while still maintaining the convenience of a score-based admission rule. This new mechanism has two key features: First, it uses the Boston (BOS) mechanism to match students with schools. Unlike the SD mechanism, which orders students according to their scores first and then allows students to choose their preferred school by following this order, the Boston mechanism asks schools to first accept students who list them as their top choices, and scores are then considered within that interested group of students. Second and importantly, the new mechanism introduces a new dimension of the mechanism design. It allows students to submit their preference list over schools before the exam is taken. Such a mechanism has in fact been used in China’s college admission system, under particular time periods and geographic regions.

The intuition for this new mechanism, which we refer to as the “pre-BOS” mechanism, in potentially implementing more ex-ante fair outcomes is simple: Under this mechanism, students have to submit their preferences based on their expected scores, which is determined by their true abilities. Those who expect lower scores have to ‘give up’ ahead of time, in applying for better schools (i.e., the more commonly favored schools) in order to secure their slot in moderately preferred schools. Slots in
the better schools are thus reserved for students who expect high scores, given that they have listed these schools as their first choices. This paper highlights the point that pre-exam preference submission may serve as a screening and commitment device, delivering the potential to improve social welfare.

The pre-BOS mechanism raises several interesting issues. First, as has been widely discussed in literature, the Boston mechanism has been regarded inferior to SD (or TTC) mechanism in the sense that it is not strategy-proof; truth telling is not an equilibrium – and it is also less likely to achieve a fair and efficient matching outcome (Abdulkadiroglu and Sonmez, 2003; Ergin and Sonmez, 2006). However, the literature has also begun to reconsider the Boston mechanism in the light of models with private information and school priority uncertainty. For example, Abdulkadiroglu, Che and Yasuda (2011) finds that if school priorities are determined by a single random tie-breaking rule, under a Bayesian equilibrium where students follow the same ordinal preference yet their cardinal preferences can be different and are private information, the matching outcome under the Boston mechanism ex-ante Pareto dominates the outcome under Deferred Acceptance (DA) mechanism. Our new mechanism introduces another source of school priority uncertainty: the uncertainty of student scores conditional on their abilities. Unlike their paper which focuses on ex-ante efficiency, this paper focuses on ex-ante fairness. Second, preference submission timing in the school choice matching problem is a new dimension of school-student matching mechanism design which is not yet fully established in literature. Although various tie-breaking rules for school priorities and private information about student preferences have been discussed (Abdulkadiroglu, et al., 2009; Edril and Ergin, 2008; Pais and Pinter, 2008; Featherstone and Niederle, 2008; Abdulkadiroglu, et al., 2011), a deliberately designed mechanism to exploit the reality of uncertain school priorities has not yet been suggested and thoroughly analyzed.

A key assumption we make in this paper is that students have homogeneous preferences on schools. Although this assumption is stringent, it is not totally unnatural. It stylizes a fact that students often have conflicts of interest with each other. Contrarily, as an extreme example, if all students prefer totally different schools, school priority rules becomes useless, students do not need any preference manipulation, and the school matching problem does not involve any welfare concerns. We find this assumption is important for us in order to deliver a tractable analysis on the pre-BOS mechanism. It also helps us to deliver some interesting results for other mechanisms studied before in literature. In our section discussion extensions, we attempt to relax this assumption to consider situations with some degree of preference heterogeneity among students, and we find that most (but not all) of our results can be extended naturally. Our paper is one of the first to explore the effects of homogeneous preferences on matching outcomes.

The paper is organized as follows: Section 2 introduces the problem with our special assumption on school priorities and student preferences, and then describes the mechanisms we will analyze. We consider the Boston Mechanism and Serial Dictatorship mechanisms, with different preference submission timing assumptions: before and after the exam. Section 3 characterizes equilibrium for mechanisms other than pre-BOS. Section 4 characterizes equilibrium under pre-BOS mechanism when ex-ante fairness is achieved. We then consider the constrained (i.e., with quota of submitted schools) pre-BOS mechanism for its possible improvement on ex-ante fairness, compared with unconstrained pre-BOS. Section 5 considers two extensions of our main results. Section 6 concludes.

2. Problem and Mechanism

2.1 The Problem

We consider a problem where N students will be matched to N schools. We assume each school
has only one slot. We consider the situation where each school has multiple slots later in the paper. The student set is denoted as \( S = \{ s_i, \ i=1, \ldots, N \} \) and the school set is denoted as \( C = \{ c_i, \ i=1, \ldots, N \} \). We assume that each student has the same strict preference ordering over schools: \( c_1 > c_2 > \ldots > c_N \). Here \( c_i > c_j \), \( i \leq j \) means that all the students prefer \( c_i \) to \( c_j \).

A matching outcome can be regarded as a function \( f: S \rightarrow C \), where each student is allocated to a school. A matching mechanism is a procedure to determine the matching outcome. In particular, each school announces in advance a priority rule for admitting students, students will report their preferences on schools (truthfully or non-truthfully), and a matching procedure (or algorithm) is used to match students with schools.

Suppose that each school uses the same priority rule of admitting students: students’ realized exam scores, denoted by \( y_i \). Each school gives higher priority to students with higher exam scores. Furthermore, each student’s realized exam score \( y_i \) is a realization of a random variable \( Y_i \), with cumulative distribution function \( \mu_i(y_i) \). Students have different intrinsic abilities so their score distributions are different. To map student ability to score distribution, we use the following definition:

**Definition (Student ability):** For any two students \( s_i \) and \( s_j \), student \( s_i \) has a higher ability than \( s_j \) if \( \mu_i(y_i) \) first order stochastically dominates \( \mu_j(y_j) \). That is, for any \( y \), \( \mu_i(y) \leq \mu_j(y) \) and there exists some \( y \), \( \mu_i(y) < \mu_j(y) \).

Without loss of generality, we assume for any students \( s_i \) and \( s_j \), \( s_i \) has a higher ability than \( s_j \) if and only if \( i < j \). The first-order stochastic domination immediately implies that \( y_i > y_j \) for any \( i < j \), where \( y = E(y) \) is a student’s expected score. It also implies that \( \Pr(y_i > y_j) > 1/2 \) for any \( i < j \). To avoid the need for a tie-breaking rule, we further assume that for any students \( i \) and \( j \), \( \Pr(y_i = y_j) = 0 \). For our future analysis, we have the following definition with regard to student score distribution.

**Definition (Competing student pair):** Two students \( s_i \) and \( s_j \) are competing with each other if there exist \( y_{i1}, y_{i2} \in \text{supp}(\mu_i), y_{j1}, y_{j2} \in \text{supp}(\mu_j) \), such that \( y_{i1} > y_{j1} \) yet \( y_{i2} < y_{j2} \).

Equivalently, two students \( i \) and \( j \) are competing with each other if \( \Pr(Y_i > Y_j) \neq 0 \) and \( \Pr(Y_j > Y_i) \neq 0 \). Intuitively, we consider two students competing with each other if their score distributions overlap.

Furthermore, suppose student scores are bounded from both below and above. Let \( y_{i \min} = \min\{y_i: y_i \in \text{supp}(\mu_i)\}, \ y_{i \max} = \max\{y_i: y_i \in \text{supp}(\mu_i)\}, \ 0 < y_{i \min} < y_{i \max} < \infty \), then two students \( s_i \) and \( s_j \), \( i < j \) are competing with each other if and only if \( y_{i \min} < y_{j \max} \). Then if two student \( s_i \) and \( s_j \), \( i < j \) are competing with each other, all the student from \( s_i \) to \( s_j \), i.e., student \( s_i, s_{i+1}, s_{i+2}, \ldots, s_j \) should be competing with each other. The proof is simple. Since for any student pair \( s_i \) and \( s_j \), \( i < j \), \( \mu_i(x) \) first order stochastically dominates \( \mu_j(x) \), we must have \( y_{i \min} > y_{j \min} \), and \( y_{i \max} > y_{j \max} \). In particular, we have \( y_{i \min} > y_{i+1 \min} > \ldots > y_{j \min} \), and \( y_{i \max} > y_{i+1 \max} > \ldots > y_{j \max} \). Since \( s_i \) and \( s_j \) are competing, we have \( y_{i \min} < y_{j \max} \). By combining these two set of inequalities, we have:

\[ y_{i \max} > y_{i+1 \max} > \ldots > y_{j \max} > y_{i \min} > y_{i+1 \min} > \ldots > y_{j \min} \]

Thus for any two students \( i_1 \) and \( i_2 \), \( i_1 < i_2 \), from the above set of inequalities, we have \( y_{i_1 \min} < y_{i_2 \max} \). They are competing with each other. Intuitively, competing relationships can only exist among "neighboring" students. Thus for all the students, we can define a symmetric situation as having a
competition degree of $k$, if competing relationship exists among any $k$ neighbored students.

**Definition (Competition degree):** A joint score distribution of all the students has a competition degree of $k$, $1 \leq k \leq N$, if competing relationship exists among all groups of at most $k$ students with neighboring intrinsic abilities, and does not exist among any group of more than $k$ students.

More concretely, if a joint score distribution has a competition degree of $k$, then for any student $s_i$, the lowest-scoring student who has a competing relationship with her is student $s_{\max \{i-k+1, 1\}}$ and the highest-scoring student who has a competing relationship with her is $s_{\min \{i+k-1, N\}}$. If $k=1$, then there is no competition relations among students at all; if $k=N$, all the students have competing relations with each other.

We now define the concepts of ex-ante fairness and ex-post fairness.

**Definition (Ex-ante Fairness).** A matching outcome $f: S \rightarrow C$ is ex-ante fair if there does not exist a pair of students $i$ and $j$ such that $\mu_i(y) < \mu_j(y)$, for any $y$, yet $f(s_i) < f(s_j)$.

**Definition (Ex-post Fairness).** A matching outcome $f: S \rightarrow C$ is ex-post fair if there does not exist a pair of students $i$ and $j$ such that $y_i > y_j$ yet $f(s_i) < f(s_j)$.

A matching outcome is ex-ante fair if students with higher intrinsic ability (thus higher expected scores) are always matched with better colleges. A matching outcome is ex-post fair if students with higher realized scores are always matched with better colleges. We denote a student $s_i$’s ex-ante fairly matched school by $f(s_i)$, and ex-post fairly matched school by $f(s_i, y)$ (for any given score realization $y=\{y_i, i=1, \ldots, N\}$). Alternatively, we can say student $s_i$ ex-ante belongs to school $c_j$ if $c_j=f(s_i)$, and ex-post belongs to school $c_j$ if $c_j=f(s_i, y)$, given the score realization $y$. It is straightforward that under our setting, $f(s_i)=c_i$.

### 2.2 Mechanisms

We will focus on four mechanisms on their possibilities of implementing ex-ante fair as well as ex-post fair matching outcome in their Nash equilibrium. We categorize mechanism from two dimensions. One is the usual division discussed in the literature, that is, Boston (BOS) mechanism versus Serial Dictatorship(SD) mechanism. The other is the newer dimension of preference submission timing, preference submission of being before exam versus after scores are known. Under the timing of preference submission before exam, a student only knows his score distribution as well as all the other students’ score distribution. Under the timing of preference submission after exam, a student has full information about all the students’ realized scores. We call the Boston mechanism with preference submission before exam the “pre-BOS” mechanism, and the Boston mechanism with preference submission after score known the “post-BOS” mechanism. “Pre-SD” and “Post-SD” mechanisms are similarly defined.

The following is the formal description of all the four mechanisms we will discuss.

---

2 Under acyclic school priority, Top Trading Cycles (TTC) mechanism is equivalent to Gale-Shapley(GS) or Deferred Acceptance(DA) Mechanism (Kesten, 2006), and SD becomes a special case of TTC mechanism (Abdulkadiroglu and Sonmez, 2003).
Pre-BOS Mechanism

Step 1. Students submit their preference order lists on all the schools.

Step 2. Exam is taken and all the students will have a realized score.

Step 3. All the students’ first ranked schools are considered. Schools will admit students who rank them first and have higher realized scores until all the slots are occupied.

Step 4. Students not admitted in previous step are considered by their second-ranked schools.

........

The procedure ends when all the students are admitted or all the ranked schools of all the students have already been considered. ■

Note that in our problem, students have no outside options (or not prefer these options), the procedure will end in step 2+N and all the students will be admitted. This is also truth for all the other mechanisms.

Pre-SD Mechanism

Step 1-2 are the same as in Pre-BOS mechanism.

Step 3. Student with the highest realized score is considered. He or she will be admitted by his or her top ranked school.

Step 4. Student with second highest realized score is considered. He or she will be admitted by his or her highest ranked school with empty slots.

........

Step 2+N. Student with the lowest realized score is considered. He or she will be admitted by his or her highest ranked school with empty slots. ■

Post-BOS Mechanism

Step 1. Exam is taken and all the students will have a realized score.

Step 2. Students submit their preference order lists on all the schools.

All the remained steps are the same as in pre-BOS mechanism. ■

Post-SD Mechanism

Step 1-2 are the same as in Post-BOS mechanism.

Step 3-(2+N) are the same as in pre-SD mechanism. ■
3. Equilibrium under Mechanisms other than pre-BOS Mechanism

We first characterize equilibrium under mechanisms other than pre-BOS. Since we assume each school follows the same priority, i.e., students’ realized score rankings, the post-BOS and post-SD mechanism will implement in its Nash equilibrium the unique ex-post fair matching outcome.

\textbf{Proposition 1:} Both Post-BOS and post-SD mechanisms will implement the unique ex-post fair matching outcome in all Nash equilibria. Furthermore, post-SD is a strategy-proof mechanism but post-BOS is not.

\textbf{Proof:} Since SD is a special case of Top Trading cycles (TTC) mechanism, all the results regarding TTC mechanism apply to SD mechanism. Our results can be derived from various theorems in Haeringer and Klijn (2009). Note also that post-BOS and post-SD is just the same as the classical BOS and SD mechanisms, with school priorities being student realized score rankings.

Since all the schools have homogeneous strict priorities, all the acyclic properties mentioned in Haeringer and Klijn (2009), (i.e., Ergin-acyclicity, Kesten-acyclicity, X-acyclicity, and strongly X-acyclicity) are satisfied. By their Theorem 7.3, the ex-post fair (or stable) matching is unique when priority structure is strongly X-acyclic. By their Theorem 7.2, TTC implements stable matching outcome in NE when priority structure is Kesten-acyclic. Since both acyclic properties are satisfied in our context, it is clear that TTC(SD) implements the unique ex-post fairness matching outcome.

By Proposition 6.1 in Haeringer and Klijn (2009), BOS implements fair matching outcome in NE, under any school priority structure. Together with the strong X-acyclicity, BOS implements the unique fair matching outcome in NE.

The strategy-proofness of post-SD/TTC mechanism has been proved in Abdulkadiroglu and Sonmez (2003), as well as the non-strategy-proofness of post-BOS mechanism.

The unique ex-post fair matching outcome is just that for any i student $s_i$ is matched with school $c_k$ if he has the kth highest score, i.e., $P(s_i, y) = c_k$, if $|\{ y_j : y_j \geq y_i \}| = k$, for any i.

Our second proposition concerns equilibrium under pre-SD mechanism.

\textbf{Proposition 2:} Pre-SD mechanism is a strategy-proof mechanism. And under its truth-telling equilibrium, the unique ex-post fair matching outcome is implemented.

\textbf{Proof:} We first prove pre-SD mechanism is strategy-proof. As we know, the usual (post-)SD mechanism is strategy-proof, i.e., truth-telling strategy is all the students’ dominant strategy (Abdulkadiroglu and Sonmez, 2003). That is to say, no matter what a student’s realized score ranking is, and other students’ realized score rankings are, truth-telling is this student’s equilibrium strategy. That is, for any realization of joint student score distribution $y = \{y_i, i = 1, \ldots, N\}$, for any student i, $u_i(T_i, S_i|y) = u_i(S_i, S_i|y)$ for any $S_i, S_i$, where $T_i$ is student $s_i$’s truth telling strategy, $S_i$ is one of student i’s strategy and $S_i$ is the combination of any other student’s strategy. So we have,

$$EU_i(T_i, S_i) = \int u_i(T_i, S_i|y)f(y)dy = \int u_i(S_i, S_i|y)f(y)dy = EU_i(S_i, S_i), \forall S_i, S_i.$$ 

Note that $EU_i(S_i, S_i)$ is just student $s_i$’s expected utility under pre-SD mechanism when the strategy
profile is \((S_i, S_i)\). So the inequality above guarantees that truth-telling is an equilibrium strategy for any student \(s_i\).

It is clear that if all the students play the truth-telling strategy, just as they are in post-SD mechanism, the matching outcome must be the same: the unique ex-post fair matching outcome.

We now characterize equilibrium outcomes of all the mechanisms other than pre-BOS: they all implement ex-post fair matching outcomes. We also characterize equilibrium strategy of pre- and post-SD mechanism: they are strategy-proof. Usually, under post-BOS mechanism, there are multiple equilibria, and equilibrium strategy is not easy to characterize. However, under our assumption of homogeneous student preference, the equilibrium strategy turns out to be easy to characterize, as shown in the following proposition:

**Proposition 3**: Under the post-BOS mechanism, any student \(s_i\) except student \(s_N\) will list their ex-post fairly-matched school as their first choice in their submitted preference list in equilibrium, and be admitted by their first choice schools.

**Proof**: According to Ergin and Sonmez (2006), and Haeringer and Klijn (2009), under the post-BOS mechanism, there must be an equilibrium where all the students list their fairly-matching schools as their top choice. By Proposition 1, we know that for each student, this fairly-matched school is unique. We then only need to prove that there is no equilibrium where students (except \(s_N\)) do not list their fairly-matched school as their top choice.

Suppose instead in an equilibrium there is one student \(s_i \neq s_N\), who does not list her fairly-matched school as her top choice. From Proposition 1, in any equilibrium, students must be matched with its unique ex-post fairly-matching schools. So student \(s_i\) is still matched with her fairly-matched school, i.e., \(c_i\). Thus student \(s_i\) is admitted by school \(c_i\) through her non-top choice, there must be an empty seat at school \(s_i\) after the first round admission.

Since \(s_i \neq s_N\), we must have another student \(s_j, j>i\), who prefers \(c_i\) to his fairly-matched school \(c_j\). Note also that under the equilibrium student \(j\) must also be admitted by her fairly-matched school \(c_j\). However, suppose now student \(s_j\) submit \(c_i\) as his top choice, then he must be admitted by \(c_i\). Thus we find a profitable deviation for student \(s_j\), which invalidates the hypothetical equilibrium.

Proposition 3 is indeed powerful. It reduces the multiple equilibria under BOS mechanism into “almost” one: except student \(s_N\), all the other students must list their equilibrium schools as their top choice. Student \(s_N\)’s choice in fact does not matter in the equilibrium, nor does other students’ non-first-choices. One implication of this proposition is that even if we restrict the number of school students can list in their preference submission, the equilibrium outcome is not affected. Indeed, this result has been found for the BOS mechanism in the general case in Haeringer and Klijn (2009) (Proposition 5.2).

Propositions 1 and 2 state that all the three mechanisms (post-BOS, pre-SD and post-SD) implement the unique ex-post fair matching outcomes. Such a matching outcome can be ex-ante fair if and only if there are no two students who are competing with each other, thus the realized score ranking is the same as the expected score ranking. This situation is one of a competition degree of 1. If the competition degree is \(k>1\), then the ex-post fair matching outcome is not ex-ante fair.

To measure the degree of non-fairness (or fairness), the concept of a blocking pair is introduced
in literature. A student-school pair in a matching outcome is a blocking pair where the student prefers the school to his allocated school, and the school admits a student with lower priority than this student or has empty seats. In our context, if the school priority rule is assumed to be the expected score (or intrinsic ability) ranking, the blocking pair is called an ex-ante blocking pair. If the priority rule is assumed to be students’ realized score ranking, the blocking pair is called an ex-post blocking pair.

It is clear that the joint distribution of student scores will fully determine the degree of ex-ante fairness under the above three mechanisms (for the pre-SD mechanism, we consider only the truth-telling equilibrium). This is because the joint distribution will determine the distribution of realized score rankings, and each realized score ranking corresponds to a unique ex-post fair matching outcome. So the realized score ranking is consistent with the expected score ranking if and only if the above three mechanisms achieve ex-ante fairness. We have the following proposition with regard to the ex-ante fairness of all the above mechanisms, under the symmetric situation where the competition degree can be defined:

**Proposition 4:** Under the pre-SD (truth-telling equilibrium), post-SD and post-BOS, suppose the competition degree is \( k \), \( 1 \leq k \leq N \). Then the maximum number of ex-ante blocking pairs in the matching outcome is:

\[
B(N, k) = \left\lfloor \frac{N}{k} \right\rfloor \cdot C(k, 2) + C(\text{mod}(N, k), 2).
\]

Where \( \left\lfloor \frac{N}{k} \right\rfloor \) represents the largest integer not exceeding \( \frac{N}{k} \), and \( \text{mod}(N, k) \) is the remainder of \( N \) divided by \( k \), and \( C(N, 2) \) represents the number of combination of choosing 2 different numbers from \( N \) numbers.

**Proof.** [Done and to be added here]

Proposition 4 highlights the point that although the three mechanisms other than pre-BOS cannot implement an ex-ante fair matching outcome, their degree of non ex-ante fairness is bounded by competition degree. Note that the maximum number of ex-ante blocking pairs \( B(N, k) \) is monotonic in \( k \). The result also indicates a way in which these three mechanisms can approach ex-ante fairness: to increase the “precision” of realized scores around expected scores, so as to reduce competition among students.

4. Implementing Ex-ante fair Matches under pre-BOS Mechanism

In this section we focus on pre-BOS mechanism, especially on its potential of implementing ex-ante fair matching outcomes in equilibrium. We first consider the pre-BOS mechanism without any restriction on the number of schools students are allowed to submit, ie. the “unconstrained pre-BOS” mechanism.

4.1. Unconstrained pre-BOS Mechanism

The following example shows a case where pre-BOS mechanism can achieve ex-ante fairness while other mechanisms cannot.

**Example 1:** Suppose there are three students \( s_1, s_2 \) and \( s_3 \), and three schools \( c_1, c_2 \) and \( c_3 \). Each school has one slot. Students have the same cardinal utility on schools as the following:
<table>
<thead>
<tr>
<th></th>
<th>School c₁</th>
<th>School c₂</th>
<th>School c₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student s₁-s₃</td>
<td>100</td>
<td>67</td>
<td>25</td>
</tr>
</tbody>
</table>

Each student has an independent score distribution as follows:

<table>
<thead>
<tr>
<th></th>
<th>Score 1 (prob. =1/2)</th>
<th>Score 2 (prob. =1/2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student s₁</td>
<td>95</td>
<td>90</td>
</tr>
<tr>
<td>Student s₂</td>
<td>94</td>
<td>89</td>
</tr>
<tr>
<td>Student s₃</td>
<td>88</td>
<td>84</td>
</tr>
</tbody>
</table>

Note that student s₁’s score first-order stochastically dominates student s₂’s score, which in turn first-order stochastically dominates student s₃’s score. In addition, student s₁ and s₂ have competing relations with each other, while student s₃ has no competition with the other two students.

Under the three mechanisms we discuss in the above section (the post-BOS, the pre- and post-SD), the matching outcome is ex-post fair. That is, the student with the highest realized score (either student s₁ or s₂) would be matched with school c₁, the student with the second highest (s₁ or s₂) score would be matched with school c₂, and the student with the lowest score (s₃) would be matched with school c₃. This matching outcome however, is not ex-ante fair. For example, if the realized score is (90, 94, 88) for student s₁-s₃, student s₂ would be matched with school c₁ and student s₁ would be matched with school c₂.

Under the pre-BOS mechanism, it is easy to characterize the following Nash equilibrium: Student s₁ would list school c₁ as his first choice, student s₂ would list school c₂ as his first choice, and student s₃ would list school c₂ as his first choice. The matching outcome is ex-ante fair. That is, student s₁ always gets school c₁, student s₂ always gets school c₂, and student s₃ always gets school c₃.

This equilibrium has some interesting properties. First, in the above characterized equilibrium, students’ second and third choices do not matter. Second, under the equilibrium student s₃ in fact is indifferent with all his strategies. This implies that there may be other equilibria in pure or mixed strategy. In fact, all the pure and mixed strategy equilibria can be characterized as the following: student s₃ play a mix strategy of (c₁, c₂, c₃) with probability q and (c₂, c₁, c₃) with a probability 1-q where q ≤ 31/42, and all the other students play the above-mentioned strategy. Third, although the game has multiple equilibria, the equilibrium outcome can be proven to be unique.

Our next proposition will characterize equilibrium under pre-BOS mechanism when the mechanism achieves ex-ante fairness. For simplicity, we consider only pure-strategy Nash equilibrium. We will discuss mixed-strategy equilibria later.

**Proposition 5:** The pre-BOS mechanism implements in its pure-strategy Nash equilibrium the ex-ante fair matching outcome only if every student except student sₙ (the student with the lowest expected score) puts his ex-ante fair matching school as his first choice, and is admitted by the first choice.

**Proof:** The ex-ante fair matching outcome implies that student sᵢ is matched to school cᵢ, for any
i. We prove that if ex-ante fair matching outcome is implemented in NE under pre-BOS mechanism, every student except student $s_N$ (the student with the lowest expected score) puts his ex-ante fair matching school as his first choice.

Suppose instead in the equilibrium where ex-ante fairness is achieved, there is one student $s_i \neq s_N$, who does not list her ex-ante fairly-matched school as her top choice. However, student $s_i$ must be matched with its unique ex-ante fairly-matched school, $c_i$, under our assumed equilibrium. Thus student $s_i$ is admitted by school $c_i$ through her non-top choice, and there must be an empty slot at school $s_i$ after the first round admission under pre-BOS mechanism.

Since $s_i \neq s_N$, we must have some student $s_j$, $j > i$, who prefers $c_i$ to his ex-ante fairly-matched school $c_j$. Note also that under the equilibrium student $s_j$ must be admitted by her fairly-matched school $c_j$. However, suppose now student $s_j$ submits $c_i$ as his top choice, then he will surely be admitted by $c_i$. Thus we find a profitable deviation for student $s_j$, which invalidates the assumed equilibrium.

Proposition 5 can also be extended to include mixed-strategy Nash equilibrium.

**Corollary 1:** The pre-BOS mechanism implements in its mixed-strategy Nash equilibrium the ex-ante fair matching outcome only if every student except student $s_N$ (the student with the lowest expected score) always puts his ex-ante fair matching school as his first choice, and is admitted by the first choice.

**Proof.** Consider a student $s_i \neq s_N$ who is playing a mixed-strategy under NE. Due to the uniqueness of ex-ante fair matching outcome, he must be admitted for sure by school $c_i$. Suppose among his support of mixed pure strategies, there is at least one pure strategy where he does not put school $c_i$ as his first choice.

Since the equilibrium outcome is ex-ante fair, there should be no other students who list school $c_i$ as his first choice in his mixed strategy. Otherwise student $s_i$ would not be admitted by school $c_i$ for sure. Now consider student $s_N$. Under the assumed equilibrium, he must be matched with school $c_N$. However, if he deviates from his equilibrium strategy to a strategy where he put school $c_i$ as his first choice, he will have positive probability of being admitting by $c_i$. Since $c_N$ is the worst school, he must be better off than under his equilibrium strategy. This again invalidates the assumed equilibrium.

In our later discussion, we will focus on pure strategy equilibrium. The reason is two-fold: First, focusing on pure strategy equilibrium facilitates our comparison between different mechanisms. As we have shown in Propositions 1-3, all the other mechanisms must have pure-strategy equilibrium. If we consider mixed-strategy equilibrium for pre-BOS, we have to consider mixed-strategy equilibrium for all the other mechanisms. Second, as we have proved in Proposition 5 and Corollary 1, although mixed strategy equilibrium may exist, under any equilibrium where ex-ante fairness is achieved, students except $s_N$ will always put their ex-ante fairly matched schools as their first choices, and be admitted by their first choice schools, no matter whether they are in a mixed-strategy or a pure strategy. So focusing on pure strategy equilibrium does not restrict our characterization of equilibrium strategy for all the students except student $s_N$. Allowing for mixed-strategy for $s_N$ will enrich our equilibrium strategy profile, although under such an equilibrium $s_N$ is indifferent for all of his mixed or pure strategy (as we will prove later).

Proposition 5 and Corollary 1 actually impose a rather strict necessary condition for pre-BOS mechanism to implement ex-ante fair matching outcome in pure or mixed strategy NE. If such a
matching outcome can really be implemented in NE, almost all the students (except student \( s_N \)) should put their ex-ante fairly-matched schools as their first choice. However, the strategy profile that almost all the students put their ex-ante fairly-matched schools as their first choice can hardly form a Nash equilibrium. Proposition 6 illustrates this point, where we denote \( u_i(c_j) \), \( 1 \leq i, j \leq N \) as the cardinal utility of school \( c_j \) for student \( s_i \):

**Proposition 6:** Pre-BOS implements ex-ante fair matching outcome in (one of) its (pure strategy) Nash equilibrium if and only if either one of the following two conditions is satisfied:

**Condition 1:** There is no competing relation between any two students.

**Condition 2:** (2.1) There exists one and only one student, \( s_k \), \( 1 < k < N \), and a subset of \( \{ s_i : 1 \leq i < k \} \), \( S_c(k) \), such that \( s_k \) has competing relationship with any student \( s_i \in S_c(k) \). No other competing relationships are allowed. (2.2) \( \max_{i \in S_c(k)} \{ \text{Prob}(y_k > y_i) * u_k(c_i) + (1 - \text{Prob}(y_k > y_i)) * u_k(c_N) \} < u_k(c_k) \), for such \( i \) and \( k \).

**Proof.**

**Sufficiency.**

Condition 1. If there is no competing relation between any two students, students’ realized score ranking is the same as their expected score ranking. The pre-BOS mechanism actually degenerates to a post-BOS mechanism, and the ex-ante fairness degenerates to ex-post fairness. According to Proposition 1, pre-BOS implements ex-ante (and ex-post) fair matching outcome in its NE.

Condition 2. We prove that a strategy profile characterized in Proposition 5, i.e., every student except student \( s_N \) (the student with the lowest expected score) puts his ex-ante fair matching school as his first choice forms a Nash equilibrium under condition 2.

We first characterize student \( s_N \)’s strategy as following: student \( s_N \) puts school \( c_k \) as his first choice.

Consider any student \( s_j \), \( 1 \leq j \leq N \), \( j \neq k \). If \( j > k \), then student \( s_j \) will have no competing relation with any other students. Given all the other students (except \( s_N \)) put their ex-ante fairly matched schools as their first choice, student \( s_j \) has no incentive to deviate from his strategy, i.e., also putting his ex-ante fairly matched schools as his first choice. Because any profitable deviation for him must consist of submitting a better school, \( c_i \), \( i < j \), as his first choice. However, since he does not have competing relations with others, he has no chance to be admitted by any better school.

Now consider student \( s_k \). He actually has competing relations with some student \( s_i \), \( i < k \). The only possibly profitable deviation for his is also putting school, \( c_i \), \( i < k \) as his first choice. If he does so, he will have two possible outcomes. Outcome 1 is that, with probability \( \text{prob}(y_k > y_i) > 0 \), he is admitted by school \( c_i \). Outcome 2 is that, with probability \( \text{prob}(y_k < y_i) > 0 \), he fails in competition with student \( s_i \), loses his first choice school, and can only be admitted by school \( c_N \). The second outcome is because all the other students except \( s_N \) have been admitted by their first choice, and student \( S_N \) has been admitted by school \( c_k \) by his first choice. So if:

\[
\max_i \{ \text{Prob}(y_k > y_i) * u_k(c_i) + (1 - \text{Prob}(y_k > y_i)) * u_k(c_N) \} < u_k(c_k), \text{ for any } i < k,
\]
Student $s_k$ will also have no incentive to deviate from the strategy stated in Proposition 5.

Finally consider student $s_N$. If all the other students have no incentive to deviate from their strategies stated in Proposition 5, he also does not have incentive to deviate. Because he has no competing relations with any other students, putting any school $c_i$, $i \neq N$ as his first choice will not give him any chance of being admitted by school $c_i$. He can only be admitted by school $c_N$ and thus is indifferent with any strategies, including the strategy we characterize for him.

**Necessity.**

According to Proposition 5, if pre-BOS mechanism implements in its pure-strategy Nash equilibrium the ex-ante fair matching outcome, then every student except student $s_N$ (the student with the lowest expected score) puts his ex-ante fair matching school as his first choice. What is left to be found is necessary conditions for such a strategy profile forming a Nash equilibrium.

Consider student $s_N$. He cannot have competing relations with any other students. Otherwise given other students’ choice as in the ex-ante fair equilibrium strategy profile, he can put a school $c_i$, $i \neq N$ as his first choice if he has competing relations with some student $s_i$. Then he can have a positive probability to be matched with school $c_i$. The matching outcome is not ex-ante fair.

Consider student $s_{N-1}$. If he does not have any competing relations with any other students, then he obviously has no incentive to deviate from the ex-ante fair equilibrium strategy profile. If he deviates, he has no chance of being admitted by any better school, since all the students except $s_N$ have put these better schools as their first choices.

Consider student $s_k$, $k=N-1, N-2, \ldots, 1$, until we find one student who does have competing relations with some student $s_i$, $i<k$. If we could not find such a student, then we have the situation of condition 1. As we proved for student $s_{N-1}$ when he has no competing relations with others, it is obvious that all the students will have no incentive to deviate from the ex-ante equilibrium strategy profile.

Now consider the situation such a student $s_k$ does exist. He should have no incentive to deviate from his ex-ante fair equilibrium strategy, i.e., putting $c_k$ as his first choice. This can happen if and only if: (1) student $s_N$ must put $s_k$ as his first choice, and (2):

$$\max_{i<k} \{ \text{Prob}(y_k>y_i) \cdot u_k(c_i) + (1-\text{Prob}(y_k>y_i)) \cdot u_k(c_N) \} \leq u_k(c_k), \text{ for any } i<k.$$  

Consider Condition (1). Suppose student $s_N$ does not put $s_k$ as his first choice. Then student $s_k$ must have incentive to deviate. Actually a deviation of putting some $c_i$ as his first choice, a school “owned” by student $s_i$ who he has competing relations with, and putting $s_k$ as his second choice will lead to a result where he is admitted by either $c_i$ or $c_k$, a better outcome than he sticks to the equilibrium strategy. However, if all the other students sticks to their equilibrium strategy, since student $s_N$ has no competing relations with others, such a strategy for $s_N$ is (one of) his equilibrium strategy. So we need no extra conditions for condition (1) to be held.

For Condition (2). Given that student $s_N$ put school $c_k$ as his first choice, this condition implies that student $s_k$ has no incentive to deviate.

The only thing left to be proved is that there cannot be another student $s_{k'}$, $1<k'<k$, who also has competing relations with some student $s_i$, $i<k'$. If such a student does exist, he must have incentive to
deviate: he can just put school \( c_i \) as his first choice and school \( c_k' \) as his second choice. He then can be admitted by either \( c_i \) or \( c_k' \), a better result than he sticks to the equilibrium strategy. \( \blacksquare \)^3

Note that our example 1 actually satisfies condition 2, with \( N=3 \), \( k=2 \), and
\[
\max_i \{\text{Prob}(y_k>y_i)u_i(c_i) + (1-\text{Prob}(y_k>y_i))(1-p_{nk})u_i(c_k) + (1-\text{Prob}(y_k>y_i))p_{nk}u_i(c_N)\} = 1/4*100 + 3/4*25 < 67.
\]

Proposition 6 actually states a result that is similar to an “impossibility” theorem for pre-BOS mechanism to implement ex-ante fair matching outcome in NE, because the competing relationship condition is so strict. Condition 1 requires that students have no competing relations with each other at all. For condition 2, the allowed competing relation is even more strict that it appears to be. Recall that we proved in section 2 that if student scores have first-order stochastic dominance relations, and any two student \( s_i \) and \( s_j \), \( i<j \) have competing relations with each other, then any two students between them, i.e., student \( s_{i1}, s_{i2}, i\leq i_1<i_2\leq j \), must have competing relations with each other. But in Proposition 6, we only allow that all the students only have competing relations with a fixed student, \( s_k \). Therefore the only competing relation we allow is that there is only a pair of students, \( s_{k-1}, s_k \), \( 1<k<N \), that have competing relations with each other. The competition degree is almost zero, if \( N \) is large.

Furthermore, under condition 1, i.e., all the students have no competing relationship with each other, actually all the four mechanism implement the same matching outcome, which is both ex-ante and ex-post fair. On the other hand, if students do have competing relations with each other, except that only two neighbored students have competing relations with each other, all the four mechanisms cannot implement ex-ante fair outcomes. Then the relevant question here is that which mechanism can implement a more ex-ante fair matching outcome, measured by the number of ex-ante blocking pairs. However, we are unable to generally compare ex-ante fairness of these four mechanisms through this view, due to its complexity. We now provide two examples to illustrate two possibilities. Example 2 shows that although pre-BOS cannot implement ex-ante fair outcomes, it can be more ex-ante fair than other mechanisms. Example 3 shows that pre-BOS mechanism can implement an outcome which is less ex-ante fair than other mechanisms.

**Example 2:** All the set-ups are the same as Example 1 except that the student score distribution changes as the following:

<table>
<thead>
<tr>
<th>Score 1 (prob. =1/2)</th>
<th>Score 2 (prob.=1/2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student ( s_1 )</td>
<td>95</td>
</tr>
<tr>
<td>Student ( s_2 )</td>
<td>91</td>
</tr>
<tr>
<td>Student ( s_3 )</td>
<td>87</td>
</tr>
</tbody>
</table>

Note that now not only Student 1 and 2 have competition relation with each other, but also Student 2, 3 have competing relations with each other.

---

3 If mixed-strategies are allowed, there can be other equilibria. Student \( s_N \) can mix several pure strategy such that in his first choice he can put a number of schools there. Suppose the set of such schools is \( C_N \). Then for any student \( s_k \), such that \( c_i \in C_N \), this student can have competing relations with some student \( s_i \), \( i<k \). The condition for student \( s_k \) not to deviate is:
\[
\max_i \{\text{Prob}(y_k>y_i)u_i(c_i) + (1-\text{Prob}(y_k>y_i))(1-p_{nk})u_i(c_k) + (1-\text{Prob}(y_k>y_i))p_{nk}u_i(c_N)\} < u_i(c_k), \text{ for any } i<k.
\]

Where \( p_{nk} \) is the probability that student \( s_N \) puts school \( c_k \) as his first choice.
Under other three mechanisms, the matching outcome is ex-post fair but not ex-ante fair. For example, if the realized score is (90, 91, 87) for Student s1-s3, then Student s2 would be matched with school c1 and Student s1 would be matched with school c2.

Under the pre-BOS mechanism, it is easy to characterize the following Nash equilibrium: Student s1 would list school c1 as his first choice, student s2 would list school c2 as his first choice, and student s3 would list school c2 as his first choice. The matching outcome is the following: student s1 will always get school c1, student s2 will get school c2 with probability 3/4 and get school c2 with probability 1/4, student s3 will get the remained school (c2 or c3).

Pre-BOS will implement a non-ex-ante fair matching outcomes when the realized score ranking (from the highest to the lowest) is (s1, s3, s2), with probability 1/4. Other mechanisms will implement a non-ex-ante fair matching outcomes when the realized score ranking is (s2, s1, s3) (with prob. 1/4) or (s1, s3, s2) (with prob. 1/4). To compare the matching outcome among four mechanisms, we can calculate their expected number of ex-ante blocking pairs as well as ex-post blocking pairs as shown in the following table:

<table>
<thead>
<tr>
<th></th>
<th>Ex-ante blocking pairs</th>
<th>Ex-post blocking pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-BOS mechanism</td>
<td>¼</td>
<td>1/4</td>
</tr>
<tr>
<td>Other mechanisms</td>
<td>½</td>
<td>0</td>
</tr>
</tbody>
</table>

So pre-BOS mechanism can implement a more ex-ante fair matching outcomes. Contrarily, other mechanisms are more likely to implement an ex-post fair matching outcomes.

Example 2 violates condition (2.1) in Proposition 6. Here student sN has competing relation with others and can “destroy” ex-ante fairness by applying for better schools. However, the best student s1 is still “protected” since condition (2.2) still holds and students who have competing relations with him (here student s2) would choose not to compete with him. How about if condition (2.2) is violated? Example 3 suggests an answer.

**Example 3:** All the set-ups are the same as in example 1 except that student cardinal preferences now change to the following table:

<table>
<thead>
<tr>
<th>School c1</th>
<th>School c2</th>
<th>School c3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student s1-s3</td>
<td>100</td>
<td>35</td>
</tr>
</tbody>
</table>

The Nash equilibrium under pre-BOS mechanism now becomes the following: student s1 and s2 put school c1 as their first choice, and student s3 puts school c2 as his first choice.

The number of ex-ante and ex-post blocking pairs under different mechanisms is shown in the following table:

<table>
<thead>
<tr>
<th></th>
<th>Ex-ante blocking pairs</th>
<th>Ex-post blocking pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-BOS mechanism</td>
<td>5/4</td>
<td>3/4</td>
</tr>
<tr>
<td>Other mechanisms</td>
<td>½</td>
<td>0</td>
</tr>
</tbody>
</table>
Pre-BOS mechanism is less likely to implement both ex-ante and ex-post fair matching outcomes than other mechanisms. ■

Example 3 violates condition (2.2) in Proposition 6. In particular, student $s_2$ has incentive to compete with student $s_1$. This violation of “ideal” equilibrium strategy “destroys” ex-ante fairness in two ways: First, it allows student $s_2$ himself to “steal” school “owned” by student $s_1$. Second, it also gives opportunity to student $s_3$, who can steal school “owned” by student $s_2$. The chain of “stealing” behaviors may heavily destroy ex-ante fairness under pre-BOS mechanism. In other words, if we can have another incentive device which can reduce students’ incentive to “risk” their first choice, we may get closer to ex-ante fairness. In the next subsection we will discuss one possibility: restricting the number of colleges students can apply for.

4.2 Constrained pre-BOS Mechanism

We now consider a constrained pre-BOS mechanism, where students can only submit one school in his preference order list. If a student is not admitted by his first choice, he will not be admitted by any school, and his utility is $u_i(0)=0$, where “0” means not being admitted at all. We further suppose $u_i(c_j)>0$, for any $c_j\neq 0$.

We can prove the following proposition.

Proposition 7: A constrained pre-BOS mechanism (where students can only apply for one school) implements in its pure-strategy Nash equilibrium ex-ante fair matching outcomes if and only if for any student $s_k$:

$$\max_i \{\text{Prob}(y_i>y_k)\ast u_k(c_i)\} \leq u_k(c_k)$$

for any $i<k$.

Proof.

Sufficiency. We prove that all the students put their ex-ante fairly matched schools as their first choice form a Nash equilibrium, if the above inequality holds.

Consider any student $s_k$, if he deviates from the assumed equilibrium strategy profile, he will either be admitted by the school he choose as his first (and only) choice, or not be admitted at all. Thus the highest pay off he can have is $\max_i \{\text{Prob}(y_i>y_k)\ast u_k(c_i)\}$. If he sticks to the assumed strategy, his payoff is $u_k(c_k)$. So if:

$$\max_i \{\text{Prob}(y_i<y_k)\ast u_k(c_i)\} \leq u_k(c_k)$$

for any $i<k$,

No student will have incentive to deviate.

Necessity. If the constrained pre-BOS mechanism (where students can only apply for one school) implements in its pure-strategy Nash equilibrium ex-ante fair matching outcomes, it must have each student put his ex-ante fairly-matched schools as their first and only choice. Under the inequality stated in the proposition, such a strategy profile also forms a Nash equilibrium. ■

Note that if conditions 1 and 2 of Proposition 6 holds, condition for Proposition 7 (the above inequality) must hold: For any student $s_k$ who has no competition relations with any student $s_i$, $i<k$, i.e.,
prob(y_i<yk)=0, thus the above inequality holds trivially. For the only student s, who has competition relations with others, condition (2.2) also implies the above inequality, since u_k(c_N)>0. But the opposite is definitely not true. In particular, unlike conditions 1 and 2 for Proposition 6, which put strict restrictions on competing relations among students, Proposition 7 does not require any competition relation restriction on students. It can hold when all the students have competing relations with each other, as long as the above inequality holds. Thus we have the following corollary:

**Corollary 2:** The constrained pre-BOS mechanism implements ex-ante fair matching outcomes in its pure-strategy Nash equilibrium if the unconstrained pre-BOS mechanism does so.

In other words, constrained pre-BOS is more likely to achieve ex-ante fairness than the unconstrained pre-BOS.

Note also that by first-order stochastic dominance, we have prob(y_i<yk)<1/2, if i<k. so if u_k(c_k)>(1/2)u_k(c_i) for any competing student pair i and k, i<k, given that u_i(0)=0, constrained pre-BOS mechanism can implement in its Nash equilibrium ex-ante fair matching outcome. This condition may be easy to satisfy because competing relations usually exist between neighbored students, so k cannot be too much larger than i, thus u_k(c_i) also cannot be too much larger than u_k(c_k).

We might be interested in matching outcome if we put such a restriction on other mechanisms. For post-BOS and SD mechanism, such a restriction will not affect matching outcomes (Haeringer and Klijn, 2009, Proposition 5.2 and theorem 5.3). If students can only submit one schools, they will just put their ex-post fair matching schools as their first choice in Nash equilibrium. For pre-SD, we can prove the same result as for pre-BOS mechanism, as stated in Corollary 3.

**Corollary 3:** A constrained pre-SD mechanism (where students can only apply for one school) implements in its pure-strategy Nash equilibrium ex-ante fair matching outcomes if and only if for any student s:

\[ \max \{ \text{Prob}(y_i>y_k) \cdot u_k(c_i) \} \leq u_k(c_k), \text{ for any } i<k. \]

The proof is almost the same as for pre-BOS so we omit it here. Proposition 7 and Corollary 3 together imply that pre-BOS and pre-SD actually converge when we restrict the submission list to contain only one school.

5. Extensions

In this section we relax some of the assumptions on our basic model. We first allow schools to have multiple slots. A similar situation we will also consider is that students have homogenous preference on schools, but their preferences may not be strict. Second, we allow students to have some degree of homogenous preferences, i.e., they have the same preferences between groups of schools, but may not be so within each group.

5.1 Multiple School Slots and Non-Strict Student Preferences

Multiple School Slots

We first consider the case that each school has multiple slots. Assume there are L schools: C={c_i: i=1,2,...,L}, with the admission quota as Q={q_i: i=1, ..., L} such that \( \Sigma q_i=N \). We denote S^i={s_i: f^i(s_i)\in c_i}, i=1, ..., L, as a student group of which each student’s ex-ante fairly matched school is c_i, and S^p={s_i: f^p(s_i,y)\in c_i} as a student group of which each student’s ex-post fairly matched school is c_i.
given the realized scores of all the students $y$.

Propositions 1-2 regarding matching outcome of mechanisms other than pre-BOS naturally extend to the multiple slot case, since they do not require assumptions on school slot. We then state a revised version of Proposition 3 on equilibrium strategy under post-BOS mechanism.

**Proposition 3E.1:** Under the post-BOS mechanism, any student $s_i$ except students ex-post belonging to $S^p_L$ will list their ex-post fairly-matched school as their first choice in their submitted preference list in equilibrium.

The proof parallels that of Proposition 3 so we omit it here.

[Possible extension of Proposition 4 here]

Proposition 5 regarding equilibrium under pre-BOS can be revised as the following:

**Proposition 5E.1:** The pre-BOS mechanism implements in its pure-strategy Nash equilibrium the ex-ante fair matching outcome only if every student except students ex-ante belonging to $S^p_L$ puts their ex-ante fair matching school as their first choice.

The proof also parallels that of Proposition 5 and is thus omitted.

Now we consider Proposition 6, which can be revised as the following:

**Proposition 6E.1:** Pre-BOS implements ex-ante fair matching outcome in (one of) its (pure strategy) Nash equilibrium if and only if either one of the following two conditions is satisfied:

**Condition 1.** There is no competing relation between any two students from different student groups $S^p_i$.

**Condition 2.** (2.1) There are $K$ students, $s_k$, $1 < k$, $f^*(s_k) \neq c_L$, and $K \leq |S^p_L|$, who have competing relationship with some student $s_i$, for any $i$ such that $f^*(s_i) > f^*(s_k)$. No other competing relationships between student groups are allowed. (2.2) $\max_i \{\text{Prob}(y_k > y_i) \cdot u_k(f^*(s_i)) + (1 - \text{Prob}(y_k > y_i)) \cdot u_k(c_L)\} < u_k(f^*(s_k))$, for any such $i$ and $k$.

The proof parallels that of Proposition 6 and is thus omitted. Note Proposition 6E.1 actually relaxes the conditions in Proposition 6. First, competing relations between students within a same student group, i.e., between students who have the same ex-ante fairly matched school, are not restricted. Second, there can be multiple students (except students ex-ante belonging to the least favored school) who have competing relations with students from other student groups, as long as the number is not too large.

Proposition 7 can be rewritten as the following:

**Proposition 7E.1:** A constrained pre-BOS mechanism (where students can only apply for one school) implements in its pure-strategy Nash equilibrium ex-ante fair matching outcomes if and only if for any student $s_k$:

$$\max_i \{\text{Prob}(y_k > y_i) \cdot u_k(f^*(s_i))\} < u_k(f^*(s_k))$$

for any $i$ and $k$ such that $f^*(s_i) > f^*(s_k)$. 18
Thus Corollary 2 still holds, as well as Corollary 3 with notational revisions.

Non-Strict Student Preferences

We now consider the case that students may have non-strict preferences on schools. We keep the assumption that each school still has one slot, and student still have homogeneous preferences. For example, students can all prefer \( c_i \) to \( c_j \), \( i < j \), but regard \( c_j \) and \( c_k \) are indifferent, for \( j < k \).

This question can be translated into a question with multiple school slots. We can redefine a set of schools which students think as indifferent as a new “aggregated” school. The difference from a multiple school slot question is, however, an equilibrium implementing ex-ante fairness under pre-BOS with multiple school slots would correspond to an equilibrium here with an additional requirement that all the students ex-ante belonging to the same “aggregated” school must put different schools within this “aggregated” school as their first choice. So a coordination problem exists here, which makes the ex-ante fairness under pre-BOS even less likely.

5.2 Some Degree of Heterogeneous Student Preferences

Our benchmark model assumes that all the students have strict homogenous preferences on schools. In section 5.1 we have relaxed this assumption a bit by allowing non-strict homogenous preferences. Here we consider another relaxation that students may have some degree of heterogeneous preferences (and some degree of heterogeneous preferences) on schools.

In particular, we partition all the \( N \) schools into a number of school groups: \( C = \bigcup_i C_i, \ i=1, \ldots, B, \ B<N, \) and \( C_i \cap C_j = \emptyset \), for any \( i \neq j \). We assume for all the students, they prefer \( c_i \) to \( c_j \) if \( c_i \in C_i, \) and \( c_j \in C_j, \) and \( i < j \). That is, students have homogenous preference on school groups. For any two schools \( c_i, c_j \) within a specific school group, if there is a student who prefers \( c_i \) to \( c_j \), there must be another student preferring \( c_j \) to \( c_i \). In other words, our partition of schools is the finest partition of schools over which all the students have the same preferences.

We can also define corresponding student group as \( S^a_i \) and \( S^p_i \), \( i=1, \ldots, B, \) with a bit of abuse of notations, such that for any student \( s_i \in S^a_i, \ P(s_i) \in C_i, \) and \( s_i \in S^p_i, \ P(s_i, y) \in C_i. \) Note that although students may have heterogeneous preferences on schools, since the acyclic assumption of school priority remains, each student still has a uniquely defined ex ant and ex-post fairly matched school.

As before, Propositions 1 and 2 do not depend on student preferences, so they are not affected. Proposition 3 can be revised as the following:

**Proposition 3E.2.** Under the post-BOS mechanism, any student \( s_i \) except students ex-post belonging to \( s_B \) will list their ex-post fairly-matched school as their first choice in their submitted preference list in equilibrium.

The proof is very similar to the proof to Proposition 3 so we omit it here.

[Possible extension of Proposition 4 here]

Proposition 5 concerning pre-BOS mechanism can be revised as the following:

**Proposition 5E.2.** The pre-BOS mechanism implements in its pure-strategy Nash equilibrium the ex-ante fair matching outcome only if every student except students ex-ante belonging to \( s_B \) puts
their ex-ante fair matching school as their first choice.

The proof is very similar with Proposition 5 thus we omit it here.

Proposition 6, however, cannot be extended (at least not easily) here. If it could be, condition 1 would be stated as: “There is no competing relations with any two students ex-ante belonging to different school groups.” Condition 2.1 would be stated as: “There are K students, sk, 1<k, f*(sk)∉CB, and K≤#(CB), who have competing relationship with some student si, for any i such that f*(si)∈Cj, j<i. No other competing relationships between student groups are allowed.” A necessary condition of (2.2) would be: “Maxi {Prob(yk>yi)*uk(f*(si))+(1-Prob(yk>yi))*uk(cj)}<uk(f*(sk)), for any such student k and i, and cj=argmin,∈CB, where c∈CB.” Here cj is defined so to consider the worst situation student sk may have when he deviates from the required equilibrium strategy. If under this situation student sk still has incentive to deviate, i.e., the inequality cannot hold, student sk must deviate. For a sufficient condition 2.1, we may redefine cj as cj=argmax,∈CB, where c∈CB.

First, it is easy to find that either condition 1 or 2 is not sufficient to guarantee that pre-BOS implement ex-ante fairness in equilibrium. Because these conditions only guarantee that students have no incentive to deviates to schools belonging to other school groups. But students may still have incentive to deviate to a school within the same ex-ante school group. The following two examples (example 4 and 5) illustrate this point.

**Example 4.** All the set-ups are the same as in example 1 except the cardinal utilities of the students become:

<table>
<thead>
<tr>
<th></th>
<th>School c1</th>
<th>School c2</th>
<th>School c3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student s1-s2</td>
<td>100</td>
<td>35</td>
<td>25</td>
</tr>
<tr>
<td>Student s3</td>
<td>35</td>
<td>100</td>
<td>25</td>
</tr>
</tbody>
</table>

So all the schools can be partitioned into 2 school groups between: C1={c1, c2} and C2={c3}. There are also no competing relations between students ex-ante belonging to these two groups, i.e., between student s1 or s2 and s3. However, there is competing relation between student s1 and s2, both being ex-ante belonging to school group C1.

The Nash equilibrium under pre-BOS mechanism is: both student s1 and s2 put school c1 as his first choice. Student s3 puts c2 as his first choice. The equilibrium outcome is not ex-ante fair.

**Example 5.** Suppose there are four student s1-s4, and four schools c1-c4. Each school has one slot. Students have the same cardinal utility on schools as the following:

<table>
<thead>
<tr>
<th></th>
<th>School c1</th>
<th>School c2</th>
<th>School c3</th>
<th>School c4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student s1-s2</td>
<td>100</td>
<td>35</td>
<td>25</td>
<td>5</td>
</tr>
<tr>
<td>Student s3</td>
<td>100</td>
<td>99</td>
<td>67</td>
<td>25</td>
</tr>
<tr>
<td>Student s4</td>
<td>35</td>
<td>100</td>
<td>25</td>
<td>5</td>
</tr>
</tbody>
</table>

Each student has an independent score distribution as the following:
Student $s_4$ has no competing relations with any students, yet students $s_1$-$s_3$ have competing relations with each other. There are three school groups: $C_1=\{c_1, c_2\}$, $C_2=\{c_3\}$, $C_3=\{c_4\}$. There is only one student ($s_1$) who have cross-group competing relations (with $s_1$ and $s_2$). For student $s_3$ we also have $(1/4)*100+(3/4)*5<67$. So condition 2 is satisfied. However, this condition could not guarantee that pre-BOS mechanism implements ex-ante fairness in NE. Suppose it could. Then we must have each student except student $s_4$ puts their ex-ante fairly matched school as their first choice (by Proposition 5E.2). Consider student $s_3$. He does so only if student $s_4$ put school $c_3$ as his first choice. But if so, both student $s_1$ and $s_2$ would put school $c_1$ as their first choice, which violates Proposition 5E.2. Ex-ante fairness should not be achieved in NE.

Second, these two conditions are also not necessary for pre-BOS mechanism to implement ex-ante fairness in NE. Remember in our necessity part of proof for Proposition 6, we derive that if pre-BOS implement ex-ante fairness, the last student $s_N$ cannot have competing relations with others. This result then is used to derive condition 1 or 2. In particular, student $s_N$ can be “used” to deter another student attempting to deviate. However, a paralleled result, i.e., students ex-ante belonging to the last school group $C_B$ cannot have competing relations with students belonging to other groups, may not be necessary for pre-BOS mechanism to implement ex-ante fairness here. Example 6 illustrates this point.

**Example 6.** All the set-ups are the same as in example 1 except the cardinal utilities of the students become:

<table>
<thead>
<tr>
<th>School $c_1$</th>
<th>School $c_2$</th>
<th>School $c_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student $s_1$</td>
<td>100</td>
<td>25</td>
</tr>
<tr>
<td>Student $s_2$-$s_3$</td>
<td>100</td>
<td>67</td>
</tr>
</tbody>
</table>

So all the schools can be partitioned into 2 school groups: $C_1=\{c_1\}$ and $C_2=\{c_2,c_3\}$. One student ex-ante belonging to $C_2$, i.e., student $s_2$ has competing relationship with student $s_1$, the student ex-ante belonging to $C_1$. However, under pre-BOS mechanism one equilibrium is: student $s_1$ puts school $c_1$ as his first choice, student $s_2$ and $s_3$ put school $c_2$ as his first choice. Ex-ante fairness is achieved.

In this example, a within-group competing relation plays a role to deter misbehavior which may destroy ex-ante fairness.

However, Proposition 7 regarding constrained pre-BOS can be extended here, as the following:

**Proposition 7E.2.** A constrained pre-BOS mechanism (where students can only apply for one school) implements in its pure-strategy Nash equilibrium ex-ante fair matching outcomes if and
only if for any student $s_k$:

$$\max_i \{\text{Prob}(y_i>y_k)u_k(f^*(s_i))\}<u_k(f^*(s_k)),$$

for any $i$ and $k$ such that $f^*(s_i)>f^*(s_k)$.

Corollary 3 with some notation revisions also holds. But Corollary 2 cannot be extended, because Proposition 6 cannot. Yet we can still conclude that constrained pre-BOS is pretty much easy to achieve ex-ante fairness, as we discuss in section 4.2.

6. Conclusions

This paper introduces a new feature of mechanism design into the school choice literature: preference submission timing. We consider a problem where school priorities are solely determined by score rankings of students from an exam taken before admission. This feature is consistent with acyclic school priority in literature. We compare two widely discussed and utilized matching procedures, the Boston and Serial Dictatorship mechanisms, interacted with two possible preference submission timings, i.e., preference submission before score is realized, and preference submission after score is realized and known.

Our discussion focuses primarily on one key assumption: that all the students have the same school preferences. It highlights a stylized fact that students usually compete with each other for the commonly regarded “better” schools. Without homogeneous school preference, this competition would be either absent or lessened. In one of our extensions, we relax this assumption by allowing some degree of preference heterogeneities, and most of our conclusions (but not all) remain valid.

To compare various mechanisms, we focus on one particular welfare property: ex-ante and ex-post fairness. Ex-ante fairness requires that students with higher intrinsic abilities (thus high expected exam scores) are matched with commonly preferred schools, while ex-post fairness requires that students with higher realized scores are matched with commonly preferred schools. Due to the uncertainty in the realization of the scores from a one-shot exam, different matching outcomes may have quite different likelihoods of satisfying different welfare criteria.

We find that among all the four mechanisms we discuss, i.e., pre-BOS (BOS with preference submission before exam), post-BOS (BOS with preference submission timing after exam), pre-SD and post-SD, three mechanisms (post-BOS, pre- and post-SD) would implement the same matching outcome in equilibrium, i.e., the ex-post fair matching outcomes. The only difference among them is strategy-proofness: post-BOS is not a truth-telling mechanism while the two SD mechanisms are. In fact, we find that under post-BOS, almost all the students (except the one with the lowest realized score) have to put their ex-post fairly matched schools as their first choice and are admitted by their first choice. This result is a direct implication of our student preference homogeneity assumption. Intuitively, due to preference homogeneity, students compete with each other so fiercely that they cannot lose their first choice admission.

When we move to the pre-BOS mechanism, we are interested in whether this “first choice admission” equilibrium remains. The answer is both “yes” and “no”. By answering “no”, we mean that although under post-BOS mechanism, ex-post fairness is always achieved under NE, under pre-BOS mechanism, ex-ante fairness is not always (and even rarely) achieved under NE. By answering “yes”, we mean that if we want to implement the ex-ante fairness under pre-BOS mechanism, we must have each student (except the one with the lowest expected score) to put their ex-ante fairly matched school as their first choices. Furthermore, the “yes” part is the reason for the “no” part. The “top-choice” equilibrium strategy profile needed for ex-ante fairness in fact hardly forms a NE. This is because if
students submit their preferences before exam, often they will perceive a chance to overpass students who have higher expected scores. Given that every else student sticks to this strategy profile, one student has good reasons to risk his first choice for a better school for which he has chance to be admitted, and ex-ante fairness thus is destroyed.

A key reason for such a risky choice or misbehavior is that students have their second (and ever more) choice as a “complete insurance” for their ex-ante fairly matched school under the suggested equilibrium. A natural way to improve the likelihood of ex-ante fairness is to remove such an insurance device. In fact, if we restrict pre-BOS mechanism by allowing students to submit only one school, ex-ante fairness is much easier to be implemented. This result is interesting if we consider the literature on constrained school choice. Haeringer and Klijn(2009) and Calsamigilia, Haeringer and Klijn (2010) find that by constraining school submission quotas, BOS mechanism can “catch up” with other mechanism because its equilibrium outcome does not depend on quotas, while those strategy-proof mechanisms (GS, TTC/SD) may be greatly affected. Our result seems to strengthen theirs in the sense that under preference submission timing of being before exam, constrained BOS mechanism can even outperform other truth-telling mechanisms with unconstrained choices.4

Our results also shed light on school choice mechanism design practice. The paper is inspired by China’s college admission, where pre-BOS mechanism had been used nationwide for more than 10 years and is still used in several provinces (e.g., Beijing and Shanghai). Although our paper does not lend overwhelmed support for this mechanism, we do recommend for improving it. First, constraints on school quotas submitted by students are necessary to deter “misbehavior” leading to ex-ante unfairness. Second, an exam which reflects intrinsic abilities more precisely will lessen competition among students and improve ex-ante fairness. Third, increasing school slots may also help, in the sense that it can lessen competition among students as well as provide students with the lowest scores more chances to deter misbehavior of other students.

One future research direction would be to relax student homogeneous preference assumption further, until we have no restriction on student preferences. Another direction would be experimental and empirical tests on those mechanisms especially the constrained pre-BOS mechanism.5 By working on these two directions we may be able to find a better mechanism to implement ex-ante fair matching outcomes. Finally, whether there exists a mechanism such that ex-ante fairness can be always implemented is still an open question.

4 Zhong, Chen and He (2004) considers the constrained pre-and post-BOS mechanism under a special case with 2 students and 2 schools, and draws a similar conclusion. They even consider a midway-BOS where students submit preference after exam is taken but before scores are known, and model it as a Bayesian game. See also Xu (2013) for an extension of this paper.

5 For some empirical and experimental studies on this topic, see Lien, Zheng and Zhong (2012), and Wu and Zhong (2012).
References


763-778.