Ex-ante Fairness in the Boston and Serial Dictatorship Mechanisms under Pre-exam and Post-exam Preference Submission

Jaimie W. Lien±, Jie Zheng† and Xiaohan Zhong‡

This Version: July 6th, 2016
Initial Version: January 20th, 2014

Abstract[1]

In a school choice mechanism, school priorities are often based on student exam scores, by which student true ability may not be perfectly revealed. An ex-post fair matching mechanism (for example, Serial Dictatorship) can be undesirable in that it is not ex-ante fair: it may not match students with higher *abilities* to better schools, although it always matches students with higher *scores* to better schools. In this paper we consider a potential way of improving ex-ante fairness - a Boston mechanism with the requirement that students submit their preferences before the exam score is realized (the "pre-BOS mechanism"). This mechanism is more likely to achieve *complete* ex-ante fairness, in that students with higher ability are always matched with better schools. However, the other mechanisms (pre-/post-SD and post-BOS) can always implement *stochastic* ex-ante fairness (students with higher ability having higher probability of admission to better schools), while pre-BOS may not.

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

JEL classification: C78, D81, I28

± Lien: Department of Decision Sciences and Managerial Economics, The Chinese University of Hong Kong, Hong Kong; jaimie.academic@gmail.com; †Zheng: Department of Economics, School of Economics and Management, Tsinghua University, Beijing, China, zhengjie@sem.tsinghua.edu.cn; ‡Zhong: Department of Economics, School of Economics and Management, Tsinghua University, Beijing, China, zhongxh@sem.tsinghua.edu.cn (corresponding author).

[1]We would like to thank the editors Yan Chen and Tom Palfrey, and an anonymous referee for their very helpful comments. We are grateful to Jingyi Xue for insights which greatly improved on the paper. For beneficial discussions, we thank Nanyang Bu, Yi-Chun Chen, Amanda Friedenberg, Ming Gao, Qian Jiao, Xinqiao Ping, John Quah, Kang Rong, Satoru Takahashi, Ning Sun, Qianfeng Tang, Ruqu Wang, Xi Weng, Chunlei Yang, Zaifu Yang, and Yongchao Zhang, as well as seminar participants at Academia Sinica (2016), Shanghai University of Finance and Economics (2014), Sun Yet-Sen University Lingnan College (2014), National University of Singapore (2014), and conference participants at Tsinghua Conference on Theoretical and Behavioral Economics, Beijing (2014). This research is funded by National Natural Science Foundation of China (#71173127, #71203112, #71303127, and #61661136002), Tsinghua University Initiative Scientific Research Grant (#20151080396 and #20151080397), Hong Kong Research Grants Council General Research Fund (#14500516) and Chinese University of Hong Kong Direct Grant. All errors are our own.

1. Introduction

In a student-to-school matching problem, a socially desirable matching result is often described as matching students with higher abilities to better schools. At least three arguments for this property can be raised: First, it is a natural extension of the widely accepted fairness concept in school choice literature (alternatively, stability, or justified-envy-freeness) originating in Gale and Shapley (1962), by further incorporating the assumption that school priorities are based on student abilities;[2] Second, it is likely to lead to higher aggregate productivity by achieving assortative matching (Becker, 1973, 1974); Third, higher ability may originate from higher effort exerted prior to the application process, thus it would seem fair to reward those who incur greater efforts, with a larger choice set which includes better schools.[3] We refer to such a matching outcome as *ex-ante fair,* or *ability-based fair*.

In practice, however, students' abilities are largely unobservable to schools and social planners. A matching mechanism attempting to achieve ex-ante fairness (or ability-based fairness) must find observable substitutes or proxies for students' abilities. Exams are one of the most frequently used proxies in education systems throughout the world: Intuitively, students with higher academic exam scores are those students with higher academic abilities. If exam scores are a perfect proxy for abilities, a serial dictatorship (SD) mechanism can be used to reach ex-ante fair matching outcomes. However, the key problem is that exam scores are often imperfect measures of students' intrinsic abilities. An ability-based fairness result might not be readily obtained under an exam- or score-based SD mechanism, which can only be what we refer to as *ex-post fair,* or *score-based fair*.

We follow in the classic mechanism design tradition of analyzing solutions for incentive compatible allocation of scarce resources with desirable welfare properties, pioneered by John O. Ledyard and others (see for example, Groves and Ledyard, 1977; Ledyard and Palfrey, 1994; and other works). In this paper, we study matching mechanisms in terms of their chances of implementing ex-ante fairness as a welfare property. Our study centers on a new mechanism, the Boston mechanism with preference submission before an exam (pre-BOS). This new mechanism has two main 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 BOS 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 further dimension in the mechanism design: It requires students to submit their preference list over schools *before* the exam is taken. We consider whether this new mechanism is able to achieve improvements in *ability-based fairness*, or *ex-ante fairness,* upon compared to other mechanisms, while still maintaining the convenience of a *score-based* admission rule.

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 must submit their preferences based on their *expected* scores, which largely reflect their true abilities. Those who expect lower scores may tend to 'give up' ahead of time on applying for better or more

[2] Schools may prefer more able students because they may establish a better reputation for schools, although this reputation may not come directly from high quality or quantity of education.

[3] In a different setting, which is commonly used in contest literature, effort may be endogenously determined and directly related to "performance", while the notion of ability can be modeled as "types", usually assumed to be exogenous. For such a framework applied to the school choice problem, see Hafalir, Hakimov, Kubler and Kurino (2015).

commonly favored schools in order to secure their slot in moderate schools. Slots in the better schools are thus reserved for students who expect higher 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 device, thus delivering the potential to improve social welfare in terms of the ex-ante fairness criterion.

The school choice matching problem, also known as the student placement problem, was first introduced and carefully studied by Balinski and Sonmez (1999), inspired by the centralized, exam-based student placement practice in Turkey. The authors evaluate the so called multi-category serial dictatorship (MCSD) mechanism, in terms of fairness, efficiency and manipulability, and prove the equivalence between the MCSD mechanism and the Gale-Shapley school optimal mechanism. Given the shortcomings in MCSD, they further propose the Gale-Shapley student optimal mechanism as the “best” mechanism that Pareto dominates all other (ex-post) fair mechanisms. Based on Balinski and Sonmez (1999)’s work, we focus on the fairness criterion, while evaluating different mechanisms, and propose new notions of fairness when there exists an imperfect correlation between each student’s ability and score.

The pre-BOS mechanism raises several interesting issues in the school choice matching literature. First, as has been widely discussed in literature, the Boston mechanism has been regarded as inferior to SD (or TTC) mechanism in the sense that it is not strategy-proof 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 light of models with private information and school priority uncertainty. For example, Abdulkadiroglu, Che and Yasuda (2011) find 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, we focus on ex- ante fairness, in other words, fairness based on students’ intrinsic abilities. Abdulkadiroglu, Che and Yasuda (2015) further consider the ex-ante efficiency properties of a variant of the deferred acceptance mechanism which allows students to express preference intensities that serve as a tie-breaker in schools’ admissions decisions. Their Choice-Augmented Deferred Acceptance (CADA) algorithm improves ex-ante efficiency through a signaling effect, similar to the screening effect of our “pre-BOS” in improving ex-ante fairness.

Second, preference submission timing in the school choice matching problem is a new dimension of the student-to-school matching mechanism design which has not been fully explored in literature. Although various tie-breaking rules for school priorities and private information about student preferences have been discussed (Abdulkadiroglu, Pathak and Roth, 2009; Erdil and Ergin, 2008; Pais and Pinter, 2008; Featherstone and Niederle, 2008; Abdulkadiroglu, Che and Yasuda, 2011), this specific source of uncertainty based on the imperfect correlation between students’ scores and abilities has not yet been rigorously analyzed from the theoretical standpoint. Some empirical studies of the effect of preference submission timing have been implemented. Wu and Zhong (2014) test the effect of school choice mechanisms with preference submission timing using field data from a top Chinese university. Lien, Zheng and Zhong (2016) test the welfare properties of the Boston and Serial Dictatorship mechanisms under pre-exam and post-exam preference submission, using laboratory experiments.[4]

Our motivating example for this line of study is China's college admissions system. As one of the world's largest centralized school choice matching systems, it sets one of its most important objectives as matching "good" students to "good" colleges, an objective which has been widely accepted as a symbol of fairness for this system. However, it has also long been criticized as only judging a student by a single one-shot exam score: the College Entrance Exam (CEE) score.[5] Also interestingly, pre-exam submission has been the dominant choice for a long time across various provinces since 1978, when the whole system was reestablished. Although post-exam submission emerged and prevailed in recent years in most provincial clearinghouses, there are still two major cities (e.g., Beijing and Shanghai) adhering to the "old" preference submission system. Meanwhile, another major policy change has also been made within this system, i.e., moving from an essentially BOS mechanism to a SD mechanism, or more precisely, a constrained SD mechanism in the sense that students are limited by the number of schools they can list on their applications. Both of these changes, from pre-exam to post-exam submission, and from BOS to SD mechanism, reinforce each other by putting more weight on this highly criticized CEE score. The welfare consequences of these policy changes are therefore highly debated and important issues to be studied. Our paper thus has real world policy relevance by considering the BOS and SD mechanisms interacted with preference submission timing (pre- or post-exam submission).

Our theoretical analysis of the ex-ante fairness properties of pre-BOS reaches a mixed conclusion. Our results do lend some support to the pre-BOS mechanism, by showing that it is more likely to achieve *complete* ex-ante fairness than other mechanisms (post-BOS, pre-SD and post-SD). That is, pre-BOS is more likely to match students with higher ability to better schools *with certainty*. Particularly when we introduce the *constrained* pre-BOS, i.e., when we limit the number of schools that can be listed by students under pre-BOS, complete ex-ante fairness becomes much easier to implement. However, pre-BOS mechanism is vulnerable in its implementation of ex-ante fairness, because it is not strategy-proof and students' equilibrium behaviors depend on their score distributions as well as their cardinal preferences. Under some conditions, pre-BOS cannot even implement *stochastic* ex-ante fairness in its equilibrium, while all the other mechanisms we consider here can always do so. Here, stochastic ex-ante fairness refers to the concept of students with higher abilities having *higher probabilities* of being matched with better schools. A policy recommendation is that a reform from pre-BOS to post-SD is reasonable given a sufficiently "precise" scoring system, yet keeping the "old" system can also be justifiable.[6]

The remainder of the paper is organized as follows: Section 2 introduces the school choice problem, establishes our assumptions on school priorities and student preferences, and then describes each of the four mechanisms we analyze. Section 3 characterizes complete ex-ante fairness under all four mechanisms. Section 4 characterizes stochastic ex-ante fairness. Section 5 considers several extensions of our main results, including multiple slots at each school, weak preferences of students, and heterogeneous student preferences. Section 6 concludes.

[4] See also Chen and Ledyard (2009) for a survey of mechanism design experiments.

[5] See Liu and Wu (2006) for a discussion of the CEE exam objectives.

[6] Here we do not discuss another possible advantage of pre-BOS, that is, ex-ante efficiency. By referring to Abdulkadiroglu, Che and Yasuda (2011), we can argue that pre-BOS can be more ex-ante efficient that other mechanisms under some conditions. For some preliminary work on this issue, see Wu and Zhong (2014), and Lien, Zheng and Zhong (2016). So as a whole, pre-BOS may still be an advantageous mechanism.

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. The student set is denoted as $S=\{s_i : i=1,...,N\}$ and the school set is denoted as $C=\{c_j : j=1,...,N\}$. We assume that each student has the same strict preference ordering over schools.[7]

Assumption 2.1 (Homogeneous strict preferences): $\forall s \in S, \forall j, j' \in \{1,...,N\}, j < j' \Rightarrow c_j \succ_s c_{j'}$.

Assumption 2.1 implies that all the students strictly prefer school c_1 to c_2, prefer c_2 to c_3, and so on.

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. Specifically, each school announces in advance a priority rule for admitting students, students then report their preferences over schools (truthfully or non-truthfully), and a matching procedure (or algorithm) is used to match students with schools.

Each school uses the same priority rule of admitting students: students' realized exam scores, denoted by y_i, $i=1,...,N$. Each school gives higher priority to students with higher exam scores. Furthermore, each student's realized exam score $y \in \left[y^{min}, y^{max}\right]$, where $\left[y^{min}, y^{max}\right] \subseteq (-\infty, \infty)$ is the score range of the exam, is a realization of an independent random variable Y_i, with a cumulative distribution function $\mu_i(y_i)$. Students have different intrinsic abilities so their score distributions are in general different. To map student ability to score distribution, we use the following definition.

Definition 2.1 (Student ability): $\forall i, i' \in \{1,...,N\}$, student s_i has a higher ability than $s_{i'}$ if $\mu_i(\cdot)$ first order stochastically dominates (FOSD) $\mu_{i'}(\cdot)$. That is, (i) $\forall y \in \left[y^{min}, y^{max}\right]$, $\mu_i(y) \le \mu_{i'}(y)$ and (ii) $\exists y' \in \left[y^{min}, y^{max}\right]$ such that $\mu_i(y') < \mu_{i'}(y')$.

It is easy to see that any ordering of students' abilities based on the definition above is transitive. However, note that such an ordering may not necessarily be complete, since it is possible that in theory there exist $i, i' \in \{1,...,N\}$ such that neither $\mu_i(\cdot)$ FOSD $\mu_{i'}(\cdot)$ nor $\mu_{i'}(\cdot)$ FOSD $\mu_i(\cdot)$. Since our main interest lies in the fairness concern when students have different abilities, it is reasonable for us to narrow down our focus to the case where the ordering is both complete and transitive throughout the paper. It is also worth noting that we assume that the ranking of students' abilities and the distributions of students' scores are commonly known to all students, and after the

[7] We will discuss the relaxations of those assumptions (i.e., single slot of schools, homogeneous student preferences) in Section 5.

exam the realized scores will also become common knowledge.[8]

Without loss of generality, $\forall i, i' \in \{1, ..., N\}$, we assume that $i < i'$ means that s_i has a higher ability than $s_{i'}$. The first-order stochastic dominance immediately implies that $\bar{y}_i > \bar{y}_{i'}$ for any $i < i'$, where $\bar{y} \equiv E(Y)$ is a student's expected score.

We denote $R(s_i, S')$ as the score ranking of student s_i in a subset of students S', where $S' \subseteq S$ and $s_i \in S'$. If $S' = S$, we simplify the notation as $R(s_i)$, or simply R_i. Obviously, a smaller number in ranking implies a higher score in the exam. We have the following property.

Lemma 2.1: For any set of n students, denoted by $S_n \equiv \{s_{i_k} : i_1 < i_2 < \cdots < i_n, k = 1, \cdots, n\}$, where $1 \le n \le N$, we have $Prob\left[R(s_{i_1}, S_n) = 1\right] > \frac{1}{n}$ and $Prob\left[R(s_{i_n}, S_n) = 1\right] < \frac{1}{n}$.

The proof is in the appendix. Lemma 2.1 implies that $Prob(Y_i \ge Y_{i'}) > \frac{1}{2}$ for any $i < i'$. To avoid discussion about tie-breaking rules, we further simplify our analysis by assuming that for any two students s_i and $s_{i'}$, it is a shy event for them to have the same score, that is, $Prob(Y_i = Y_{i'} = y) = 0, \forall y \in \left[y^{min}, y^{max}\right]$. Thus we have $Prob(Y_i > Y_{i'}) > \frac{1}{2}$ since $Prob(Y_i = Y_{i'}) = 0$.

As our analysis will show, the following definition, *pair of competing students*, is a key concept in deriving our results.

Definition 2.2 (Pair of competing students): Two students s_i and $s_{i'}$ are *competing* with each other if $Prob(Y_i > Y_{i'}) \in (0,1)$ or $Prob(Y_i < Y_{i'}) \in (0,1)$.

Note that the definition above describes a competing relationship between two students under a very general setting of joint score distributions. Since we rule out the case of having a tie ($Prob(Y_i = Y_{i'}) = 0$), we have $Prob(Y_i > Y_{i'}) \in (0,1)$ if and only if $Prob(Y_i < Y_{i'}) \in (0,1)$. Furthermore, since in our framework Y_i and $Y_{i'}$ are independent for any $i, i' \in \{1, ..., N\}$, the definition of the pair of competing students is equivalent to the following one.

Definition 2.2' (Pair of competing students): Two students s_i and $s_{i'}$ are *competing* with each other if there exist $y_i, y_i' \in supp(\mu_i)$ and $y_{i'}, y_{i'}' \in supp(\mu_{i'})$ such that $y_i > y_{i'}$ yet $y_i' < y_{i'}'$.

In a special case where supports of score distributions are intervals, the condition for a pair of

[8] Score uncertainties affect the equilibrium only through the expected payoffs in our complete information game setup. However in reality, abilities as well as score distributions may be private information. It is also plausible that students may not know their own true abilities and receive private noisy signals about their abilities. In those contexts, a Bayesian game modeling framework must be introduced to handle the analysis. We leave this to our future research. One point we may hypothesize is that when there are a large amount of students, the Bayesian game analysis may lead to conclusions similar to those from our complete information game framework.

students competing with each other becomes $supp(\mu_i) \cap supp(\mu_{i'}) \neq \emptyset$. Intuitively, in this case, two students are competing with each other if and only if their score distributions overlap.

Furthermore, suppose student scores are bounded from both below and above. $\forall i = 1, ..., N$, let $y_i^{inf} \equiv inf\{y_i : y_i \in supp(\mu_i)\}$ and $y_i^{sup} \equiv sup\{y_i : y_i \in supp(\mu_i)\}$, where $y^{min} \leq y_i^{inf} \leq y_i^{sup} \leq y^{max}$. Thus two students s_i and $s_{i'}$ with $i < i'$ are competing with each other if and only if $y_i^{inf} < y_{i'}^{sup}$. We have the following lemma concerning the competing relationship among students.

Lemma 2.2: Suppose that student scores have first-order stochastic dominance relations. Given $1 \leq i < i' \leq N$, if two students s_i and $s_{i'}$ are competing with each other, then for any two students $s_{i''}$ and $s_{i'''}$ with $i \leq i'' < i''' \leq i'$, they are competing with each other.

The proof is in the appendix. Intuitively, competing relationships exist among all the "neighboring" students. For all the students, we can define a symmetric situation as having a *competition degree of* n, if competing relationship exists among any n neighboring students.

Definition 2.3 (Competition degree): A joint score distribution of all the students has a competition degree of n, where $1 \leq n \leq N$, if competing relationship exists within any group of at most n students with neighboring intrinsic abilities, and does not exist within any group of more than n students.

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

2.2 Ex-ante and Ex-post Fairness

In general, our concept of fairness reflects that "better" students should go to "better" schools. This can be defined rigorously as follows, from both ex-ante (ability-based) and ex-post (score-based) perspectives.

Definition 2.4 (Complete Ex-ante Fairness): A matching outcome $f : S \to C$ is *completely* ex-ante fair if for any pair of students s_i and $s_{i'}$ with $i < i'$, we have $f(s_i) \succ_{s_i} f(s_{i'})$.

The concept of "complete" ex-ante fairness is the ex-ante analogy to the existing concept in the literature of ex-post fairness. However, in the ex-ante case, there is a potentially useful weaker notion of fairness which only requires probabilistic fairness rather than one with certainty. This is represented by our concept of stochastic ex-ante fairness. Let p_{ij} denote the probability of student s_i matched with j th most preferred school in the matching outcome, and $p_i(k) \equiv \sum_{j=1}^{k} p_{ij}, k = 1, \dots, N$ be student s_i's corresponding cumulative distribution function over the schools. The concept of stochastic ex-ante fairness can be defined as follows.

Definition 2.5 (Stochastic Ex-ante Fairness): A matching outcome $f: S \to C$ is *stochastically* ex-ante fair if for any pair of students s_i and $s_{i'}$ with $i < i'$, we have $p_{i'}(\cdot)$ FOSD $p_i(\cdot)$, which implies (i) $\forall k \in \{1, \dots, N\}, p_i(k) \geq p_{i'}(k)$, and (ii) $\exists k \in \{1, \dots, N\}, p_i(k) > p_{i'}(k)$.[9]

Finally we give the definition for ex-post fairness:

Definition 2.6 (Ex-post Fairness): A matching outcome $f: S \to C$ is ex-post fair if for any score realization $\boldsymbol{y} = \{y_{i''} : i'' = 1, \cdots, N\}$, for any pair of students s_i and $s_{i'}$ with $y_i > y_{i'}$ (or equivalently, $R_i < R_{i'}$), we have $f(s_i) \succ_{s_i} f(s_{i'})$.

Note that complete ex-ante fairness implies for any student s_i, $p_{ii} = 1$. Therefore, a matching outcome is *completely* ex-ante fair if students with higher intrinsic ability (thus higher expected scores) are *always* matched with better colleges. A matching outcome is *stochastically* ex-ante fair if students with higher intrinsic ability are *more likely* to be matched with better colleges, in the sense that the "distribution of school quality" of a better student FOSD that of a worse student. It is clear that complete ex-ante fairness implies stochastic ex-ante fairness, thus the former can be regarded as a "first-best" criterion and the latter a "second-best" one. Our analysis later on will show that the distinguishing between these two ex-ante concepts is critical for our result. A matching outcome is (completely) ex-post fair if students with higher *realized scores* are *always* matched with better colleges. We denote a student s_i's ex-ante fair matched school by $f^a(s_i)$, and the ex-post fair matched school by $f^p(s_i, \boldsymbol{y})$, given the score realization $\boldsymbol{y} = \{y_{i'} : i' = 1, \cdots, N\}$. Alternatively, we can say that student s_i ex-ante belongs to school c_j if $c_j = f^a(s_i)$, and ex-post belongs to school c_j if $c_j = f^p(s_i, \boldsymbol{y})$, given the score realization $\boldsymbol{y}$. It is straightforward that under our setting $f^a(s_i) = c_i$.

A key assumption we make in this paper is that students have homogeneous preferences over schools. Although this assumption is stringent, it is not totally artificial. It represents a stylized fact that students have conflicts of interest with each other. Contrarily, as an extreme example, if all students prefer totally different schools, school priority rules will become useless, students do not need any preference manipulation, and the school matching problem does not involve any welfare concerns. We find that this assumption is important for us in order to deliver a tractable analysis on the pre-BOS mechanism, and also helps us to deliver some interesting results for other mechanisms studied in literature. It also largely reflects the reality in China's college admissions system where students' preferences over colleges are at least similar to some degree. In Section 5, which discusses extensions to the model, we attempt to relax this assumption by considering situations with some degree of preference heterogeneity among students, and we find that most (but not all) of our results can be extended naturally.

2.3 Mechanisms

We study four mechanisms on their possibilities of implementing ex-ante (completely or stochastically) fair matching outcome in their Nash equilibria. We categorize mechanisms from two

---

[9] Note that our labeling of students' preferences over schools has more preferred schools denoted by smaller numbers, which is the opposite of the standard labeling in the definition of FOSD. Thus, $p_{i'}(\cdot)$ FOSD $p_i(\cdot)$.

dimensions. One is the usual division discussed in the literature, that is, Boston (BOS) mechanism versus Serial Dictatorship (SD) mechanism.[10] The other dimension is the new feature of preference submission timing: preference submission occurring before the exam is taken versus after the scores are known. Under the timing of preference submission before the exam, a student only knows his score distribution as well as all the other students' score distributions. Under the timing of preference submission after the exam, a student has full information about all students' realized scores. We call the Boston mechanism with preference submission before the exam the "pre-BOS" mechanism, and the Boston mechanism with preference submission after the exam with known scores the "post-BOS" mechanism. "Pre-SD" and "post-SD" mechanisms are similarly defined.

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

Pre-BOS Mechanism

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

Step 2. An 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 the previous step are considered by their second-ranked schools.

…….

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

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

Pre-SD Mechanism

Steps 1 and 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 the second highest realized score is considered. He or she will be admitted by his or her highest ranked school that has an empty slot.

……

Step N+2. Student with the lowest realized score is considered. He or she will be admitted by his or her highest ranked school that has an empty slot. □

[10] 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).

Post-BOS Mechanism

Step 1. An exam is taken and every student will have a realized score.

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

All the remaining steps are the same as in pre-BOS mechanism. □

Post-SD Mechanism

Steps 1 and 2 are the same as in post-BOS mechanism.

Steps 3 to (N+2) are the same as in pre-SD mechanism. □

3. Implementing Complete Ex-ante Fairness

We first focus on the complete ex-ante fairness issue, by considering under what conditions each of the four mechanisms can implement such a socially-desirable matching outcome.

3.1 Mechanisms other than pre-BOS

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 their Nash equilibrium the unique ex-post fair matching outcome.

Proposition 3.1: Both the 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.

The proof is in the appendix.[11] The unique ex-post fair matching outcome is merely that for any $i = 1, ..., N$, student s_i is matched with school c_j if s_i has the jth highest score, i.e., $f^P(s_i, \boldsymbol{y}) = c_j$, if $R_i = j$ for any $\boldsymbol{y}$.

Our second proposition concerns the equilibrium under the pre-SD mechanism.

Proposition 3.2: The pre-SD mechanism is a strategy-proof mechanism, and in its truth-telling equilibrium the unique ex-post fair matching outcome is implemented.

---

[11] Balinski and Sonmez (1999) shows that the SD mechanism can be characterized as the unique (ex-post) fair mechanism that is Pareto optimal (and it is also strategy-proof), which constitutes part of our result in Proposition 3.1. Since our results regarding the BOS mechanism require different proof techniques, to be consistent, we adopt a different approach for the proof as a whole.

The proof is in the appendix. Propositions 3.1 and 3.2 state that all the three mechanisms (post-BOS, pre-SD and post-SD) implement the unique ex-post fair matching outcome.

Such a matching outcome can be ex-ante fair if and only if there are no competing relationships between any two students. That is, for any two students s_i and $s_{i'}$, $i < i'$ must imply $Prob(Y_i > Y_{i'}) = 1$ and $Prob(Y_i < Y_{i'}) = 0$. Otherwise, if $Prob(Y_i < Y_{i'}) > 0$, then $Prob(R_i > R_{i'}) > 0$, and by the definition of ex-post fairness, this implies $Prob(f^p(s_{i'}, \boldsymbol{Y}) \succ_s f^p(s_i, \boldsymbol{Y})) > 0$. Note that this result also holds the other way around. We describe this finding in the following theorem:

Theorem 3.1: The pre-SD (under its truth telling equilibrium), post-BOS and post-SD mechanisms implement complete ex-ante fairness if and only if students have no competing relationship with each other.

3.2 Pre-BOS Mechanism

In this subsection we focus on the pre-BOS mechanism, especially on its potential of implementing completely ex-ante fair matching outcomes in equilibrium. The following example, in which students have competing relationships, shows a case where the pre-BOS mechanism can achieve complete ex-ante fairness while the other three 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 utilities on schools as the following:

	School c_1	School c_2	School c_3
Student s_1-s_3	100	67	25

Each student has an independent score distribution as follows:

	Score 1(prob. =1/2)	Score 2 (prob.=1/2)
Student s_1	95	90
Student s_2	94	89
Student s_3	88	84

Note that student s_1's score first-order stochastically dominates student s_2's score, which in turn first-order stochastically dominates student s_3's score. In addition, students s_1 and s_2 have competing relations with each other, while student s_3 has no competition with the other two students.

Under the three mechanisms we discuss in the above section (post-BOS, pre-SD and post-SD), the matching outcome is ex-post fair. That is, the student with the highest realized score (either student s_1 or s_2) would be matched with school c_1, the student with the second highest score (s_1 or s_2) would be matched with school c_2, and the student with the lowest score (s_3) would be matched with school c_3. This matching outcome however, is not completely ex-ante fair. To see this, suppose that the realized scores are (90, 94, 88) for students s_1, s_2 and s_3, then student s_2 would be matched with school c_1 and student s_1 would be matched with school c_2.

Under the pre-BOS mechanism, it is easy to characterize the following Nash equilibrium: student s_1 listing school c_1 as his first choice, s_2 listing c_2 as his first choice, and s_3 listing c_2 as his first choice. The matching outcome is completely ex-ante fair. That is, s_i is always matched with c_i, for $i = 1, 2, 3$.[12]

The random matching outcome for each mechanism is characterized by $\{p_{ij} : i, j \in \{1, 2, 3\}\}$, and shown in the following table. Part (3) of the table also shows the differences of p_{ij}'s between different mechanisms. It is clear that pre-BOS implements complete ex-ante fairness while the others are not. □

(1) pre-/post-SD, post-BOS

p_{ij}	c_1	c_2	c_3
s_1	3/4	1/4	0
s_2	1/4	3/4	0
s_3	0	0	1

(2) pre-BOS

p_{ij}	c_1	c_2	c_3
s_1	1	0	0
s_2	0	1	0
s_3	0	0	1

(3) Diff. (2)-(1)

Δp_{ij}	c_1	c_2	c_3
s_1	1/4	-1/4	0
s_2	-1/4	1/4	0
s_3	0	0	0

Our next proposition characterizes the equilibrium under the pre-BOS mechanism *when the mechanism achieves complete ex-ante fairness*.

[12] This equilibrium has some interesting properties. First, in the above characterized equilibrium, students' second and third choices do not matter. Second, in equilibrium student s_3 is in fact indifferent between all his strategies. This implies that there may be other equilibria in pure or mixed strategies. In fact, all the pure and mixed strategy equilibria can be characterized as follows: Student s_3 plays a mixed strategy of (c_1, c_2, c_3) with probability q and (c_2, c_1, c_3) with probability $1-q$ where $q < \frac{31}{42}$, and all the other students play the above-mentioned strategies. If s_3 chose a higher probability (i.e., $q \geq \frac{31}{42}$) of playing (c_1, c_2, c_3), then it would be optimal for student s_2 to choose (c_1, c_2, c_3) instead of his equilibrium strategy, to compete with s_1 for c_1, due to the lack of threat from s_3 of stealing c_2. However, if s_2 chose (c_1, c_2, c_3), s_3 would find it optimal to choose c_2 as his first choice, contradicting the assumption that $q \geq \frac{31}{42}$. Third, although the game has multiple equilibria, the equilibrium outcome can be proven to be unique.

Proposition 3.3: The pre-BOS mechanism implements in its pure-strategy Nash equilibrium the completely ex-ante fair matching outcome only if every student except s_N (the least able student) puts his ex-ante fair matched school as his first choice, and is admitted by that school.

The proof is in the appendix. Proposition 3.3 can also be extended when we consider mixed-strategies.

Corollary 3.1 (of Proposition 3.3): The pre-BOS mechanism implements in its mixed-strategy Nash equilibrium the completely ex-ante fair matching outcome only if every student except s_N (the least able student) always puts his ex-ante fair matched school as his first choice, and is admitted by that school.

The proof is in the Appendix. 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 3.1 and 3.2, all the other mechanisms must have a 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 3.3 and Corollary 3.1, although mixed-strategy equilibrium may exist, under any equilibrium where complete ex-ante fairness is achieved, students except s_N will always put their ex-ante fair matched schools as their first choices, and be admitted by those schools, no matter whether they are in a mixed-strategy or a pure-strategy equilibrium. 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 strategies (as we will prove later).

Proposition 3.3 and Corollary 3.1 actually impose a rather strict necessary condition for the pre-BOS mechanism to implement completely 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 fair matched schools as their first choices. However, the strategy profile where almost all the students put their ex-ante fair matched schools as their first choices can hardly form a Nash equilibrium under the pre-BOS mechanism. Theorem 3.2 illustrates this point, where we use $u_i(c_j)$ to denote the cardinal utility of getting admitted by school c_j for student s_i, where $i, j = 1, \cdots, N$.

Theorem 3.2: The pre-BOS mechanism implements completely 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 relationship between any two students.

Condition 2: There exist a unique student s_k, where $1 < k < N$, and a nonempty subset of students $S_k \subseteq \{s_i : 1 \le i < k\}$, such that

(2.1) $\forall \tilde{i}, \hat{i} \in \{1, \cdots, N\}$, $s_{\tilde{i}}$ and $s_{\hat{i}}$ are a pair of competing students if and only if

$\exists s_i \in S_k$ ***satisfying*** $\{\tilde{i},\hat{i}\}=\{i,k\}$ **;**

$$\textbf{(2.2)}\ \underset{s_i\in S_k}{Max}\ \ \{Prob(Y_k>Y_i)\cdot u_k(c_i)+(1-Prob(Y_k>Y_i))\cdot u_k(c_N)\}\le u_k(c_k)\,.$$

Proof:

Sufficiency:

Condition 1. If there is no competing relationship 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 complete ex-ante fairness degenerates to ex-post fairness. According to Proposition 3.1, pre-BOS implements completely ex-ante (and ex-post) fair matching outcome in its NE.

Condition 2. We would like to prove that the strategy profile, where every student except s_N (the student with the lowest expected score) puts his ex-ante fair matching school as his first choice and student s_N puts school c_k as his first choice, forms a Nash equilibrium under condition 2,

First consider any student $s_{i'}$, where $1\le i'<N$ and $i'\ne k$. *Case 1:* $s_{i'}\notin S_k$: In this case student $s_{i'}$ has no competing relationship with any other student. *Case 2:* $s_{i'}\in S_k$: In this case student $s_{i'}$ has no competing relationship with any students who are more able than $s_{i'}$. Given that all the other students (except s_N) put their ex-ante fair matched schools as their first choices, student $s_{i'}$ has no incentive to deviate from his strategy, i.e., also putting his ex-ante fair matched school as his first choice. Because any profitable deviation for him must consist of submitting a better school, c_i, $i<i'$, as his first choice. However, since he does not have competing relationship with any students who are more able than him, he has no chance to be admitted by any better school.

Then consider student s_k. He actually has competing relationship with all the students in S_k. The only possibly profitable deviation for him is to put some school, c_i, where i is chosen such that $s_i\in S_k$, as his first choice. If he does so, he will have two possible outcomes. Outcome 1 is that, with probability $Prob(Y_k>Y_i)$, he is admitted by school c_i. Outcome 2 is that, with probability $1-Prob(Y_k>Y_i)$, 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 due to the fact that all other students except s_N have been admitted by their first choices, and student s_N has been admitted by school c_k as his first choice. So if condition (2.2) holds, that is

$$\underset{s_i\in S_k}{Max}\ \ \{Prob(Y_k>Y_i)\cdot u_k(c_i)+(1-Prob(Y_k>Y_i))\cdot u_k(c_N)\}\le u_k(c_k)\,,$$

student s_k will also have no incentive to deviate from the strategy stated in Proposition 3.3.

Finally consider student s_N. If all the other students have no incentive to deviate from their strategies stated in Proposition 3.3, s_N also does not have incentive to deviate: Since he has no competing relationship with any other students, putting any school c_i with $i \neq N$ 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 of listing school c_k as his first choice.

Necessity:

According to Proposition 3.3, if the pre-BOS mechanism implements in its pure-strategy Nash equilibrium the completely ex-ante fair matching outcome, then every student except s_N (the student with the lowest expected score) puts his ex-ante fair matched school as his first choice. What is left to be found is the necessary condition for such a strategy profile to form a Nash equilibrium.

First, consider student s_N. He cannot have competing relationship with any other students. Otherwise given other students' choices as in the completely ex-ante fair equilibrium strategy profile, he can put a school c_i with $i < N$ as his first choice if he has competing relationship with some student s_i. By doing so he can have a positive probability to be matched with school c_i. Such a matching outcome is not completely ex-ante fair.

Then, consider student s_{N-1}. If he does not have any competing relationship with any other students, then he obviously has no incentive to deviate from the completely 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.

Sequentially check student s_k, $k = N-1, N-2, \cdots, 1$, until we find one student who does have competing relationship with some student s_i with $i < k$. If we could not find such a student, then we have the situation described by Condition 1. As we have shown for student s_{N-1} when he has no competing relationship with others, it is obvious that all the students will have no incentive to deviate from the completely ex-ante equilibrium strategy profile.

Now consider the situation where such a student s_k with $1 < k < N$ does exist. Denote the set of students that s_k has competing relationship with by S_k. Obviously for any student $s_i \in S_k$, we have $1 \leq i < k$. We would like to make sure that s_k has no incentive to deviate from his completely ex-ante fair equilibrium strategy, i.e., putting c_k as his first choice. This can be guaranteed if and only if (i) student s_N must put c_k as his first choice, and (ii)

$$\underset{s_i \in S_k}{Max} \left\{ Prob(Y_k > Y_i) \cdot u_k(c_i) + (1 - Prob(Y_k > Y_i)) \cdot u_k(c_N) \right\} \leq u_k(c_k) .$$

Consider Condition (i) first. Suppose that student s_N does not put c_k as his first choice. Then student s_k must have incentive to deviate. Actually a deviation of putting some c_i such that $s_i \in S_k$ as his first choice, a school "owned" by student s_i who he has competing relationship with, and putting c_k as his second choice will lead to a result where he is admitted by either c_i or c_k, which is a better result than the one where he sticks to the equilibrium strategy. However, if all the other students sticks to their equilibrium strategy, since student s_N has no competing relationship with others, such a strategy for s_N is (one of) his equilibrium strategies.

For Condition (ii). Given that student s_N puts school c_k as his first choice, this condition guarantees that student s_k has no incentive to deviate.

The only thing left to be proved is that there cannot be another student $s_{k'}$ with $1 < k' < k$ who also has competing relationship with some student s_i with $1 \le i < k'$. 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 can then be admitted by either c_i or $c_{k'}$, which is a better result than the one where he sticks to the equilibrium strategy.□[13]

Note that our example 1 actually satisfies Condition 2, with $N = 3$, $k = 2$, $S_2 = \{s_1\}$ and $\underset{i \in S_k}{Max} \ \{Prob(Y_k > Y_i) \cdot u_k(c_i) + (1 - Prob(Y_k > Y_i)) \cdot u_k(c_N)\} = \frac{175}{4} < 67 = u_k(c_k)$.

Compared with Theorem 3.1, Theorem 3.2 implies that pre-BOS indeed is more likely to implement complete ex-ante fairness than other mechanisms: It can implement complete ex-ante fairness even if students have some competing relationship with each other (by Condition 2). However, the superiority of pre-BOS is marginal, because the competing relationship allowed is very strict. Recall that if student scores have first-order stochastic dominance relations, Lemma 2.2 shows that if two students s_i and $s_{i'}$ are competing with each other, then for any two students $s_{i''}$ and $s_{i'''}$ with $i \le i'' < i''' \le i'$, they are also competing with each other. However, in Theorem 3.2, we only allow that a subset of students have competing relations with a unique student, s_k. Therefore the only competing relationship we allow for is that there is only one pair of competing students, s_{k-1} and s_k, where $1 < k < N$. The competition degree is almost zero, if N is large.

In our extension section later on we are going to relax some of our assumptions here. It turns out by those relaxations, the pre-BOS mechanism can have a larger chance to “defeat” the three other mechanisms. Relaxation includes multiple school slots, non-strict student preferences, as well as some degree of heterogeneous student preferences. An even more significant improvement on pre-BOS for implementing complete ex-ante fairness is by using a “constrained” pre-BOS mechanism, which we analyze below.

---

[13] If mixed-strategies are allowed, there can be other equilibria. Student s_N can mix several pure strategies 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 $c_k \in C_N$, this student can have competing relations with a subset of students $S_k \subseteq \{s_i : 1 \le i < k\}$. The condition for student s_k not to deviate is: $\underset{s_i \in S_k}{Max} \ \{Prob(Y_k > Y_i) \cdot u_k(c_i) + (1 - Prob(Y_k > Y_i)) \cdot [(1 - p_N^k) \cdot u_k(c_k) + p_N^k \cdot u_k(c_N)]\} \le u_k(c_k)$, where p_N^k is the probability with which student s_N puts school c_k as his first choice.

3.3 Constrained pre-BOS Mechanism

Under the unconstrained pre-exam BOS mechanism, if other students put their ex-ante fair matched schools as their first choices, a student s_i who has competing relationship with a better student will have a strong incentive to deviate by competing with such a better student. The reason is that he is insured by his second choice: If student s_N does not put c_i as his first choice, s_i can put c_i as his second choice and being admitted by c_i if he fails in the competition. One way to weaken this competing (or stealing) incentive is to get rid of this second choice.

We now consider a constrained pre-BOS mechanism, where each student is allowed to submit only one school in his preference ordering list. If a student is not admitted by his first choice, he will not be admitted by any school, and his utility is $u_i(\varnothing)=0$, where "$\varnothing$" means not being admitted at all. We further assume that $u_i(c_j)\geq 0$, for any $i,j=1,\cdots,N$.

We might be interested in the matching outcome if we put such a restriction on other mechanisms. It is worth noting that when each student's preference submission is only allowed to contain one school, the constrained pre-BOS mechanism as specified above is equivalent to the analogously defined constrained pre-SD mechanism, while more general versions of these two mechanisms differ from each other. The same relationship holds for the constrained post-BOS and post-SD mechanisms. For these two mechanisms, such a restriction will not affect matching outcomes (Haeringer and Klijn, 2009, Proposition 5.2 and Theorem 5.3), and if every student can only submit one school, the students will just put their ex-post fair matched schools as their first choices in Nash equilibrium.

By putting a limit on the number of schools students are allowed to list, we have the following proposition.

Proposition 3.4: The constrained pre-BOS (or equivalently, constrained pre-SD) mechanism (where each student can only apply for one school) implements in its pure-strategy Nash equilibrium completely ex-ante fair matching outcomes if and only if for any student s_k, for any student s_i such that s_i and s_k are a pair of competing students and $i<k$,

$$\boldsymbol{Prob(Y_k>Y_i)\cdot u_k(c_i)\leq u_k(c_k)}.$$

The proof is in the Appendix. It is worth mentioning that by FOSD relations of student scores, we have $Prob(Y_k>Y_i)<\frac{1}{2}$, if $i<k$. So if $u_k(c_k)\geq\frac{1}{2}u_k(c_i)$ for any pair of competing students s_i and s_k with $i<k$, given that $u_i(\varnothing)=0$, the constrained pre-BOS/SD mechanism can implement in its Nash equilibrium completely 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 $u_k(c_k)$.

Note also that if Condition 1 or 2 of Theorem 3.2 holds, the condition for Proposition 3.4 must hold: If there is no competing relationship between any two students (Condition 1), the condition for Proposition 3.4 holds trivially; For the case where only one student s_k has competing relationship with others (Condition 2), condition (2.2) also implies the above inequality, since $u_k(c_N) \geq 0$. However, the opposite is definitely not true. In particular, unlike Conditions 1 and 2 for Theorem 3.2, which put strict restrictions on competing relationship among students, Proposition 3.4 does not require any restriction on the competing relationship among students. It can hold even when all the students have competing relationship with each other, as long as the above inequality holds. Thus we have the following corollary.

Corollary 3.2 (of Proposition 3.4): The constrained pre-BOS/SD mechanism implements complete ex-ante fair matching outcomes in its pure-strategy Nash equilibrium if the unconstrained pre-BOS mechanism does so.

In other words, the constrained pre-BOS is more likely to achieve complete ex-ante fairness than the unconstrained pre-BOS. This result is interesting if we consider the literature on constrained school choice. Haeringer and Klijn (2009) and Calsamiglia, Haeringer and Klijn (2010) find that by limiting school submission quota, the BOS mechanism can "catch up" with other mechanisms because its equilibrium outcome does not depends 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 the preference submission timing of before the exam, the *constrained* BOS mechanism can even outperform other truth-telling mechanisms with *unconstrained* choices.[14] This also parallels the finding in Abdulkadiroglu, Che and Yasuda (2015) regarding their CADA algorithm that a richer message space for students to express their preferences need not yield superior ex-ante welfare outcomes.

4. Implementing Stochastic Ex-ante Fairness

Our previous results favor the pre-BOS mechanism than the other three mechanisms (pre-/post-SD, post-BOS), by showing that pre-BOS (especially the constrained one) is more likely to implement complete ex-ante fairness. Is it our policy recommendation that we should adopt the (constrained) pre-BOS mechanism to improve ex-ante fairness? Not necessarily. Note that the superiority of pre-BOS (either constrained or unconstrained) is somehow vulnerable, because it depends on the cardinal utilities of students. This is in turn because of two reasons: (i) Uncertainty of scores when students submit their preference list, and (ii) Non-strategy-proofness of the BOS mechanism. Lack of any of these two elements may make cardinal preference considerations unnecessary. This suggests that for the other three mechanisms, although they are less likely to implement the complete ex-ante fairness, they may be more robust to implement some kinds of ex-ante fairness, which turns out to be the stochastic ex-ante fairness (see Definition 2.5). In this section we will consider stochastic ex-ante fairness issues for all these four mechanisms (pre-/post-BOS/SD).

[14] Zhong, Chen and He (2004) considers the constrained pre-and post-BOS mechanisms under a special case with 2 students and 2 schools, and draws a similar conclusion. They also 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 Xu (2013) for an extension of that paper.

4.1 Mechanisms other than Pre-BOS

For the three mechanisms other than pre-BOS (i.e., post-BOS, pre-/post-SD), we have the following theorem.

Theorem 4.1: Pre-SD (in its truth-telling equilibrium), post-SD and post-BOS implement stochastic ex-ante fairness in their (pure-strategy) Nash equilibrium.

The proof (in the appendix) is straightforward after showing that the FOSD relations for the score distribution implies the reversed FOSD relations for the score ranking (Lemma A.1 in the appendix).

Although these three mechanisms can always implement stochastic ex-ante fairness, the degree of ex-ante fairness within the family of stochastic ex-ante fairness can vary a lot. On one extreme case, complete ex-ante fairness is stochastically ex-ante fair. On the other extreme case, each student having the same and equal probability of being admitted by each school is also very close to (although not the same as) stochastic ex-ante fairness. Intuitively, the degree of ex-ante fairness within this stochastic ex-ante fairness depends on how precisely student scores reflect their true abilities.

4.2 Pre-BOS Mechanism

Now we consider the pre-BOS mechanism. We first provide two examples: Example 2 shows that pre-BOS can be *more* stochastically ex-ante fair than the others; while Example 3 shows that pre-BOS can implement an outcome that is not stochastically ex-ante fair, in which case pre-BOS is obviously less stochastically ex-ante fair than the others.

Example 2: All the set-ups are the same as in Example 1 except that the students' score distributions are modified as the following:

	Score 1(prob. =1/2)	Score 2 (prob.=1/2)
Student s_1	95	90
Student s_2	91	86
Student s_3	87	82

Note that now not only students s_1 and s_2 have competing relationship with each other, but also students s_2 and s_3 have competing relationship with each other. (By Definition 2.3, the competition degree is 2 here.)

Under other three mechanisms, the matching outcome is ex-post fair: Student with the highest realized score will get the best school, and so on. The equilibrium matching outcome $\{p_{ij} : i, j \in \{1,2,3\}\}$ is shown in the table below (part (1)). It is easy to verify that this matching outcome is stochastically ex-ante fair.

Under the pre-BOS mechanism, it is easy to characterize the following Nash equilibrium:

student s_1 listing school c_1 as his first choice, s_2 listing c_2 as his first choice, and s_3 listing c_2 as his first choice. The equilibrium matching outcome $\{p_{ij} : i, j \in \{1,2,3\}\}$ is shown in the table below (part (2)). This matching result is also stochastically ex-ante fair.

It is easy to see that the pre-BOS implement *more* stochastic ex-ante fairness than other mechanisms. This is shown in part (3) in the following table, by calculating the differences of p_{ij}'s between the two matching outcome for each i and j. Pre-BOS always weakly increases the probabilities of ex-ante fair matching (i.e., $\{p_{ii} : i \in \{1,2,3\}\}$) and weakly decrease probabilities of ex-ante unfair matching (i.e, $\{p_{ij} : i, j \in \{1,2,3\}, i \neq j,\}$), compared with other mechanisms. □

(1) pre-/post-SD, post-BOS

p_{ij}	c_1	c_2	c_3
s_1	3/4	1/4	0
s_2	1/4	1/2	1/4
s_3	0	1/4	3/4

(2) pre-BOS

p_{ij}	c_1	c_2	c_3
s_1	1	0	0
s_2	0	3/4	1/4
s_3	0	1/4	3/4

(3) Diff. (2)-(1)

Δp_{ij}	c_1	c_2	c_3
s_1	1/4	-1/4	0
s_2	-1/4	1/4	0
s_3	0	0	0

In Example 2, under the pre-BOS mechanism, Condition (2.1) in Theorem 3.2 is violated: Student $s_N \equiv s_3$ has competing relationship with others and can "destroy" ex-ante fairness by applying for better schools. However, the best student s_1 is still "protected" since Condition (2.2) still holds and students who have competing relationship with him (here student s_2) would choose not to compete with him. This is the main reason why in this example pre-BOS is more stochastically ex-ante fair than the other mechanisms. It also seems paradoxical that student s_1 is protected just because student s_2 is not protected – it faces the (successful) competition from student s_3.

Example 2 seems to suggest that the violation of Condition (2.1) may not have "huge" damaging effect on (stochastic) ex-ante fairness. It can even be helpful for ex-ante fairness by deterring "stealing" behavior of some more able students. However, the violation of Condition (2.2) may be more severe, since the direct effect of this would be to encourage the "stealing" behavior. Example 3 suggests that this could be true.

Example 3: All the set-ups are the same as in Example 2 except that the students' cardinal preferences are now modified as in the following table:

	School c_1	School c_2	School c_3

Student s_1-s_3	100	35	25

The Nash equilibrium under the pre-BOS mechanism now becomes the following: students s_1 and s_2 listing school c_1 as their first choice, and s_3 listing c_2 as his first choice. The equilibrium under the other three mechanisms is irrelevant to cardinal preferences and thus is the same as in Example 2.

The random matching outcome under different mechanisms are shown in the following table. It can be verified that the matching outcome under pre-BOS is no longer stochastically ex-ante fair. In particular, $p_{21}+p_{22}=\frac{1}{4}<1=p_{31}+p_{32}$. It can also be argued that pre-BOS is less stochastically ex-ante fair than the other three mechanisms: For students s_2 and s_3, it increases the probability of "mismatch" and decreases the probability of ex-ante fair matching; For student s_1, it does not affect the probability of ex-ante fair matching, but it increases the probability of a more severe "mismatch" (p_{13}) and decreases that of a less severe "mismatch" (p_{12}). □

(1) pre-/post-SD, post-BOS

p_{ij}	c_1	c_2	c_3
s_1	3/4	1/4	0
s_2	1/4	1/2	1/4
s_3	0	1/4	3/4

(2) pre-BOS

p_{ij}	c_1	c_2	c_3
s_1	3/4	0	1/4
s_2	1/4	0	3/4
s_3	0	1	0

(3) Diff. (2)-(1)

Δp_{ij}	c_1	c_2	c_3
s_1	0	-1/4	1/4
s_2	0	-1/2	1/2
s_3	0	3/4	-3/4

Example 3 violates Condition (2.2) in Theorem 3.2. 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" the school "owned" by student s_1. Second, it also creates an opportunity for student s_3, who can steal the school "owned" by student s_2. The chain of "stealing" behaviors may heavily destroy ex-ante fairness under the pre-BOS mechanism.

It turns out difficult to fully characterize stochastic ex-ante fairness under the pre-BOS mechanism. Unlike pre-SD and post-SD, pre-BOS is not strategy-proof and students have incentive and space to manipulate their reported preferences. Also unlike post-BOS, where score uncertainty is resolved and students can only consider their ordinal preferences (at least in pure-strategy equilibrium), under pre-BOS equilibrium behaviors are jointly determined by score distributions and cardinal preferences. Just as our Examples 1-3 illustrate, any change in score distribution and cardinal preferences may result in changes in the equilibrium matching outcomes for pre-BOS.

However, we are still able to characterize a necessary condition under pre-BOS for stochastic ex-ante fairness (or a sufficient condition for stochastic ex-ante *unfairness*) under certain scenario with simple score distributions, defined in Definition 2.3, i.e., a symmetric competing relationship among students.

Proposition 4.1: Suppose that the student score joint distribution satisfies n-degree ($n > 1$) competing relationship (as in Definition 2.3), then pre-BOS will implement stochastic ex-ante fairness in its weakly-dominant pure-strategy NE only if $N \le \frac{n(n+1)}{2}$.

The proof is in the appendix.

Proposition 4.1 states that if the competition degree is not large enough, stochastic ex-ante fairness cannot be implemented under pre-BOS. However, if competition degree is $n = 1$, i.e., there is no competing relationship among student, as our Theorem 3.2 suggests, *complete* ex-ante fairness can be implemented under pre-BOS. The idea behind Proposition 4.1 is the following: If any competition relationship is allowed such that the competing (or "stealing") behavior of students cannot be deterred (when Condition 2 of Theorem 3.2 cannot hold), to maintain a stochastic ex-ante fairness, competition must be so fierce that those who steal other students' schools must also face "stealing" behavior from the other students, and so on. Look back to our Examples 2 and 3, which satisfies the symmetric competing relation with $N = 3$ and $n = 2$ so that $N = \frac{n(n+1)}{2}$. There may exist a weakly-dominant pure-strategy NE such that stochastic ex-ante fairness can be implemented. Example 2 indeed shows the case (while Example 3 do not).

An interesting special case would be $n = N$. That is, all the students have competing relationship with each other. Under this condition, we can show that for some cardinal student preferences (but not all), truth-telling can be a (symmetric) equilibrium.[15] Therefore, in such cases pre-BOS can be equally stochastically ex-ante fair as other mechanisms.

Consider the following numeric example which is constructed to capture the reality of China's college admissions. In each year, roughly 9 million students attended China's college entrance examination. Therefore for each of 30 provinces and each track (humanity or science), on average there are 150,000 students, which can be considered as the number of students set in a single matching problem. Assuming the existence of a n-degree competing relationship among all the students, for stochastic ex-ante fairness to be possible, we need $150000 = N \le \frac{n(n+1)}{2}$, or $n \ge 548$, or $\frac{n}{N} > 0.365\%$. The full score in the exam is typically 750. Suppose that a typical student has a uniform score distribution with a span of $2b$ and an average of x. That is, his possible score is equally likely drawn from $[x-b, x+b]$. For any two students with average scores of x_1 and x_2, with $x_1 < x_2$, if they want to have competing relationship with each other, the largest possible gap of their average scores is $x_2 - x_1 = 2b$. Suppose that the average score x for all students are uniformly distributed from score 150 to 750. (150 is roughly the cutoff points for those colleges at the bottom level.) So we need $2b = x_2 - x_1 \ge 0.365\% \cdot (750 - 150) = 2.192$, or $b \ge 1.096$. This almost surely holds in reality.

[15] Furthermore, if $n < N$, truth-telling is not an equilibrium for any cardinal utilities.

5. Extensions

In this section we relax some of the assumptions on our basic model. First, we allow schools to have multiple slots. A similar situation we will also consider is that students have homogenous preferences over 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.

Our analysis here will focus on the pre-BOS mechanism. For the other mechanisms, results do not depend on whether those assumptions hold or not, except for the strict student preferences assumption, which is usually assumed in literature. We will see that this assumption is similar to the multiple school slots assumption, so it would be easy to address.

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 that there are L schools $C=\{c_j : j=1,...,L\}$, with the admission quota as $Q=\{q_j : j=1,...,L\}$ such that $\sum_{j=1}^{L} q_j = N$. For any $j=1,\cdots,L$, we use $S_j^a \equiv \{s_i \in S : f^a(s_i)=c_j\}$ to denote the *ex-ante student group of* c_j in which each student's ex-ante fair matched school is c_j, and $S_j^p \equiv \{s_i \in S : f^p(s_i, \boldsymbol{y})=c_j\}$ to denote the *ex-post student group of* c_j in which each student's ex-post fair matched school is c_j given the realized scores of all the students $\boldsymbol{y}=\{y_i : i=1,\cdots,N\}$.

Proposition 3.3 regarding the necessary condition for complete ex-ante fairness in equilibrium under pre-BOS can be revised as the following:

Proposition 5.1: The pre-BOS mechanism implements in its pure-strategy Nash equilibrium the completely ex-ante fair matching outcome only if every student except students in S_L^a puts their ex-ante fair matched school as their first choice.

The proof parallels that of Proposition 3.3 and is thus omitted.

Now we consider Theorem 3.2, the necessary and sufficient condition for complete ex-ante fairness under pre-BOS, which can be revised as the following:

Proposition 5.2: The pre-BOS mechanism 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 relationship between any two students from different S_j^a, $j=1,\cdots,L$.

Condition 2: There exist a subset of students $S^M \subseteq \{s_k \in S : 1<k \le N-q_L\}$ with

$\#(S^M) \le q_L$, ***and for any*** $s_k \in S^M$ ***there exist a nonempty subset of students*** $S_k^M \subseteq \left\{s_i : f^\alpha(s_i) \succ_s f^\alpha(s_k)\right\}$***, such that***

(2.1) $\forall \tilde{i}, \hat{i} \in \left\{1,\cdots,N\right\}$, $s_{\tilde{i}}$ ***and*** $s_{\hat{i}}$ ***are a pair of competing students if and only if*** $f^\alpha(s_{\tilde{i}}) = f^\alpha(s_{\hat{i}})$ ***or*** $\exists s_k \in S^M, \exists s_i \in S_k^M$ ***satisfying*** $\left\{\tilde{i}, \hat{i}\right\} = \left\{i, k\right\}$***;***

(2.2) $\forall s_k \in S^M$,

$$\underset{s_i \in S_k^M}{Max} \left\{Prob(Y_k > Y_i) \cdot u_k(f^a(s_i)) + (1 - Prob(Y_k > Y_i)) \cdot u_k(c_L)\right\} \le u_k(f^a(s_k)).$$

The proof parallels that of Theorem 3.2 and is thus omitted. Note that Proposition 5.2 actually relaxes the conditions in Theorem 3.2. First, competing relationship between students within the same ex-ante student group, i.e., between students who have the same ex-ante fair matched school, are not restricted. Second, there can be multiple students (except students in S_L^a) who have competing relationship with students from other ex-ante student groups, as long as the number of these students is not too large ($\#(S^M) \le \#(S_L^a)$).

Proposition 3.4, characterizing the sufficient and necessary condition for constrained pre-BOS/SD to implement complete ex-ante fairness, can be rewritten as the following:

Proposition 5.3: The constrained pre-BOS/SD mechanism (where each student can only apply for one school) implements in its pure-strategy Nash equilibrium completely ex-ante fair matching outcomes if and only if for any student s_k***, for any student*** s_i ***such that*** s_i ***and*** s_k ***are a pair of competing students and*** $f^\alpha(s_i) \succ_s f^\alpha(s_k)$***,***

$$Prob(Y_k > Y_i) \cdot u_k(f^\alpha(s_i)) \le u_k(f^\alpha(s_k)).$$

It is easy to see that Corollary 3.2 still holds.

Proposition 4.1, which states a necessary condition for pre-BOS to implement stochastic ex-ante fairness cannot be extended, because the symmetric competition relationship that held before may not hold any more in the case of multiple school slots: Students who have competing relationship with each other but belong to the same ex-ante student group do not "compete" with each other anymore; yet students who have competing relationship with each other but do not belong to the same ex-ante student group still have "real" competing relationship.

Non-Strict Student Preferences

We now consider the case that students may have non-strict preferences over 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_j to c_{j+1}, but are indifferent between c_{j+1} and c_{j+2}.

This question can be translated into one 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, that 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 in the same ex-ante student group of that "aggregated" school must put *different* schools within this "aggregated" school as their first choices. So a coordination problem exists here, which makes the ex-ante fairness under any mechanism less likely.

5.2 Some Degree of Heterogeneous Student Preferences

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

In particular, we partition all the N schools into a number of school groups: $C = \bigcup_{m=1}^{B} C_m$ and $C_m \bigcap C_{m'} = \varnothing$, $\forall m \neq m'$ where $1 < B < N$ and $m, m' = 1, \cdots, B$. We assume that all the students prefer c_j to $c_{j'}$ if $c_j \in C_m$, $c_{j'} \in C_{m'}$ and $m < m'$. That is, students have homogenous preference over school groups. For any two schools c_j and $c_{j'}$ within a specific school group, if there is a student who prefers c_j to $c_{j'}$, there must be another student who prefers $c_{j'}$ to c_j. 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 the corresponding ex-ante and ex-post student groups as S_m^a and S_m^p, $m = 1, \cdots, B$, with a bit of abuse of notations, such that for any student $s_i \in S_m^a$, we have $f^a(s_i) \in C_m$, and for any student $s_i \in S_m^p$, we have $f^p(s_i, \boldsymbol{y}) \in C_m$. 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-ante and ex-post fair matched school.

Proposition 3.3 regarding the necessary condition for the pre-BOS mechanism to implement complete ex-ante fairness can be revised as the following:

Proposition 5.4: The pre-BOS mechanism implements in its pure-strategy Nash equilibrium the completely ex-ante fair matching outcome only if every student except students in S_B^a puts their ex-ante fair matched school as their first choice.

The proof is very similar to that of Proposition 3.3, and thus we omit it here.

Theorem 3.2, however, cannot be extended (at least not in a straightforward way) here. If it could be extended, Condition 1 would be stated as: "There is no competing relationship between any two students from different S_k^a, $k = 1, \cdots, B$." If possible, a plausible Condition 2 would be stated as: "There exist a subset of students $S^H \subseteq \left\{ s_k \in S : 1 < k \leq N - \#(S_B^a) \right\}$ with $\#(S^H) \leq \#(S_B^a)$, and for any $s_k \in S^H$ there exist a nonempty subset of students $S_k^H \subseteq \left\{ s_i : f^\alpha(s_i) \succ_s f^\alpha(s_k) \right\}$, such that: (2.1) $\forall \tilde{i}, \hat{i} \in \left\{ 1, \cdots, N \right\}$, $s_{\tilde{i}}$ and $s_{\hat{i}}$ are a pair of competing students if and only if $\exists m \in \left\{ 1, \cdots, B \right\}$ satisfying $s_{\tilde{i}}, s_{\hat{i}} \in S_m^a$ or $\exists s_k \in S^H$, $\exists s_i \in S_k^H$ satisfying $\left\{ \tilde{i}, \hat{i} \right\} = \left\{ i, k \right\}$; (2.2) $\forall s_k \in S^H$, $\underset{s_i \in S_k^H}{Max} \left\{ Prob(Y_k > Y_i) \cdot u_k(f^a(s_i)) + (1 - Prob(Y_k > Y_i)) \cdot u_k(c_j(k)) \right\} \leq u_k(f^a(s_k))$." Here, for a weaker version of (2.2) $c_j(k)$ is defined as $c_j(k) = \underset{c \in C_B}{argmin} \quad u_k(c)$ to consider the worst situation student s_k may have when he deviates from the required equilibrium strategy. If under this situation student s_k still has incentive to deviate, i.e., the inequality cannot hold, student s_k must deviate. For a stronger version of (2.2), we may redefine $c_j(k)$ as $c_j(k) = \underset{c \in C_B}{argmax} \quad u_k(c)$.

First, it is easy to find that either "pseudo-"Condition 1 or 2 is not sufficient to guarantee that pre-BOS implements ex-ante fairness in equilibrium. Because these conditions only guarantee that students have no incentive to deviate to schools in the other school groups. But students may still have incentive to deviate to a school within the same ex-ante school group. The following example illustrates this point.

Example 4: All the set-ups are the same as in Example 1 except that the cardinal utilities of the students are modified as in the following table:

	School c_1	School c_2	School c_3
Student s_1-s_2	100	35	25
Student s_3	35	100	25

So all the schools can be partitioned into 2 school groups: $C_1 = \{c_1, c_2\}$ and $C_2 = \{c_3\}$. There are also no competing relationship between students ex-ante belonging to these two groups, i.e., between student $\{s_1, s_2\}$ and s_3. However, there is competing relationship between student s_1 and s_2, both being ex-ante belonging to school group C_1.

The Nash equilibrium under the pre-BOS mechanism is: Both students s_1 and s_2 listing school c_1 as their first choice and student s_3 listing c_2 as his first choice. The equilibrium matching outcome is not completely ex-ante fair. □

Second, these two conditions are also not necessary for the pre-BOS mechanism to implement ex-ante fairness in NE. Recall that in the necessity part of proof for Theorem 3.2, we have derived that if pre-BOS implements ex-ante fairness, the last student s_N cannot have competing relationship with others. This result is then 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 relationship with students belonging to other groups, may not be necessary for the pre-BOS mechanism to implement ex-ante fairness here. Example 5 illustrates this point.

Example 5: All the set-ups are the same as in Example 1 except that the cardinal utilities of the students are modified as in the following table:

	School c_1	School c_2	School c_3
Student s_1	100	25	67
Student s_2-s_3	100	67	25

Now 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 the pre-BOS mechanism an equilibrium is: Student s_1 listing school c_1 as his first choice, and both students s_2 and s_3 listing school c_2 as their first choice. Complete ex-ante fairness is achieved. □

In this example, a within-group competing relationship plays a role to deter the "stealing" behavior which may destroy ex-ante fairness.

However, Proposition 3.4 regarding the constrained pre-BOS/SD mechanism can be extended here, as the following:

Proposition 5.5: The constrained pre-BOS/SD mechanism (where each student 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, for any student s_i such that s_i and s_k are a pair of competing students and $f^{\alpha}(s_i) \succ_s f^{\alpha}(s_k)$,

$$\boldsymbol{Prob(Y_k > Y_i) \cdot u_k(f^{\alpha}(s_i)) \leq u_k(f^{\alpha}(s_k))}.$$

When we allow for some degree of heterogeneous preferences by the students, Corollary 3.2 cannot be extended, simply because Theorem 3.2 cannot be extended. However, we can still conclude that constrained pre-BOS/SD mechanism is pretty much easy to achieve ex-ante fairness, as we have discussed in Section 3.3.

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 the literature. We compare two widely discussed and implemented matching procedures, the Boston and Serial Dictatorship mechanisms, interacted with two possible preference submission timings, i.e., preference submission before the exam score is realized, and preference submission after the exam score is realized and known.

To compare various mechanisms, we focus on one particular welfare property: ex-ante fairness (or ability-based fairness). We consider two forms of ex-ante fairness: Complete ex-ante fairness requires that students with higher intrinsic abilities (thus high expected exam scores) are matched with commonly preferred schools with certainty; Stochastic ex-ante fairness requires that students with higher abilities have higher probabilities of being matched with commonly preferred schools.

Among all the four mechanisms we study, i.e., pre-BOS (BOS with preference submission before exam), post-BOS (BOS with preference submission after exam), pre-SD (SD with preference submission before exam), and post-SD (SD with preference submission after exam), we find the following results. First, three mechanisms (post-BOS, pre-SD and post-SD) will implement the same matching outcome in equilibrium, i.e., the stochastically ex-ante fair matching outcomes. The only difference among them is in students' incentives to reveal their true preferences: Post-BOS is not a strategy-proof mechanism while the two SD mechanisms are. Second, pre-BOS is more likely to implement complete ex-ante fairness, especially when we limit the number of schools each student can list. However, pre-BOS is more "vulnerable" than other mechanisms in the sense that it will implement even non-stochastically ex-ante fair outcomes with positive probability.

Our results lend some explanatory power to China's college admissions system. The difference in results under the two different concepts of ex-ante fairness helps us to better understand why pre-BOS and other mechanisms coexist in the current college admissions system, in different provinces. In particular, pre-BOS is not as "bad" as perhaps originally thought, especially when we consider a constrained pre-BOS, which is the version of the mechanism that occurs in practice.

Based on our study, the policy implications are the following: First, if the current system is a pre-SD, post-SD or post-BOS mechanism, the best way to improve ex-ante fairness is to enhance its examination system so that scores can become a better proxy for abilities.[16] Second, if the current system is a pre-BOS mechanism, we may want to improve ex-ante fairness by imposing limits on the number of schools submitted by students in their application forms, or alternatively, increasing school slots or enlarging the sets of "equally preferred" schools to ease competition among students. However, we still need to be cautious of the possibility of "bad" results under this non-robust mechanism.

Our results have heavily relied on one key assumption: All the students have the same preferences over schools. In our extensions, we relax this assumption by allowing some degree of preference heterogeneity. A future research direction would be to relax this assumption further, in an ideal case of no restriction on student preferences. Another research direction would be experimental and empirical tests on those mechanisms especially the constrained pre-BOS mechanism.[17] By pursuing these two directions we may be able to find a better mechanism to implement ex-ante fair matching outcomes. Although our framework has assumed common knowledge of students' abilities and score distributions, future work may consider the issues we discuss in this paper in a private information, Bayesian game framework. Our model has made a first attempt to model the relationship between students' abilities and their score performances, by assuming an exogenous but stochastic relation between these two factors. Future work can pursue an endogenous relationship between ability and score, along the lines of Hafalir, Hakimov, Kubler and Kurino (2015). Finally, whether there exists a mechanism in which complete ex-ante fairness can be *always* implemented, is still an open question.

References

[1] Abdulkadiroglu, Atila, Yeon-Koo Che, and Yosuke Yasuda, Resolving Conflicting Preferences in

[16] Increasing score precision may still help pre-BOS implement ex-ante fairness *if competing relationship among students is completely erased.* Otherwise, by Proposition 4.1, it may even decrease the possibility of ex-ante fairness under pre-BOS, by decreasing n.

[17] For some empirical and experimental studies on this topic, see Lien, Zheng and Zhong (2016), and Wu and Zhong (2014).

School Choice: the “Boston” Mechanism Reconsidered, *American Economic Review,* 101(Feb., 2011), pp. 399-410.

[2] Abdulkadiroglu, Atila, Yeon-Koo Che, and Yosuke Yasuda, Expanding “Choice” in School Choice, *American Economic Journal: Microeconomics*, 7:1 (Feb 2015), pp. 1 – 42.

[3] Abdulkadiroglu, Atila, Parag A. Pathak, and Alvin E. Roth, Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match, *American Economic Review*, 99:5 (Dec., 2009), pp. 1954–1978.

[4] Abdulkadiroglu, Atlia, and Tayfun Sonmez, School Choice: A Mechanism Design Approach, *American Economic Review*, 93:3 (Jun., 2003), pp. 729-747.

[5] Balinski, Michel and Tayfun Sonmez, A Tale of Two Mechanisms: Student Placement, *Journal of Economic Theory,* 84 (1999), pp. 73 – 94.

[6] Becker, Gary S., A Theory of Marriage, Part I, *Journal of Political Economy*, 81:4 (Jul. - Aug., 1973), pp. 813–846.

[7] Becker, Gary S., A Theory of Marriage, Part II, *Journal of Political Economy*, 82: 2 (Mar. - Apr., 1974), pp. S11-S26.

[8] Calsamiglia, Caterina, Guillaume Haeringer and Flip Klijn, Constrained School Choice: An Experimental Study, *American Economic Review,* 100:4 (Sep. 2010), pp. 1860-1874.

[9] Chen, Yan and John O. Ledyard, “Market Design Experiments”, in The New Palgrave Dictionary of Economics, 2nd edition (2009).

[10] Erdil, Aytek, and Haluk Ergin, What’s the Matter with Tie-Breaking? Improving Efficiency in School Choice, *American Economic Review*, 98:3 (Jun., 2008), pp. 669-689.

[11] Ergin, Haluk, and Tayfun Sonmez, Games of School Choice under the Boston Mechanism, *Journal of Public Economics*, 90(2006), pp. 215-237.

[12] Featherstone, Clayton, and Muriel Niederle, Ex-ante Efficiency in School Choice Mechanisms: An Experimental Investigation, *mimeo*, 2008.

[13] Gale, D., and L. S. Sharpley, College Admissions and the Stability of Marriage, *American Mathematical Monthly*, 69:1 (Jan., 1962), pp. 9-15.

[14] Groves, Theodore and John Ledyard, Optimal Allocation of Public Goods: A Solution to the “Free Rider” Problem, *Econometrica*, 45:4 (1977), p. 783 – 810.

[15] Haeringer, Guillaume, and Flip Klijn, Constrained School Choice, *Journal of Economic Theory*, 144(2009), pp. 1921-1947.

[16] Hafalir, Isa E., Rustamdjan Hakimov, Dorothea Kubler and Morimitsu Kurino, College Admissions with Entrance Exams: Centralized versus Decentralized, *Working Paper,* December 2015.

[17] Kesten, Onur, On Two Competing Mechanisms for Priority-based Allocation Problems, *Journal of Economic Theory,* 127(2006), pp. 155-171.

[18] Ledyard, John O., and Thomas R. Palfrey, Voting and Lottery Drafts as Efficient Public Goods Mechanisms, *The Review of Economic Studies*, 61:2 (1994), p. 327 – 355.

[19] Lien, Jaimie W., Jie Zheng, and Xiaohan Zhong, Preference Submission Timing in School Choice Matching: Testing Fairness and Efficiency in the Laboratory, *Experimental Economics*, 19:1(Mar., 2016), pp. 116-150.

[20] Liu, Haifeng, and Wu Qiong, "Consequences of college entrance exams in China and the reform challenges", *KEDI Journal of Educational Policy*, Vol. 3 (2006), No. 1, p. 7 – 21.

[21] Pais, Joana, and Agnes Pinter, School Choice and Information: An Experimental Study on Matching Mechanisms, *Games and Economic Behavior*, 64 (2008), pp. 303-328.

[22] Wu, Binzhen, and Xiaohan Zhong, Matching Mechanisms and Matching quality: Evidence from a Top University in China, *Games and Economic Behavior*, 84 (Mar. 2014) , pp.196–215

[23] Xu, Boli, A Game Theoretical Model and Social Welfare Analysis on Preference Submission in China's College Admission, mimeo (in Chinese), 2013.

[24] Zhong, Xiaohan, Na Chen and Yunfan He, Where Have All the Flowers Gone: An Analysis of the National College Entrance Exam, *China Economic Quarterly (in Chinese)*, 3:3 (Apr. 2004), pp. 763-778.

Appendix: Proofs

Proof of Lemma 2.1

We have the following two inequalities:

$$Prob\left[R\left(s_{i_1},S_n\right)=1\right]=Prob\left[Y_{i_1}\geq Y_{i_k},\forall k=2,\cdots,n\right]=\prod_{k=2}^{n}Prob\left[Y_{i_1}\geq Y_{i_k}\right]=\int_{y^{min}}^{y^{max}}\prod_{k=2}^{n}Prob\left[y\geq Y_{i_k}\right]d\mu_{i_1}(y)$$

$$=\int_{y^{min}}^{y^{max}}\prod_{k=2}^{n}\mu_{i_k}(y)d\mu_{i_1}(y)>\int_{y^{min}}^{y^{max}}\left[\mu_{i_1}(y)\right]^{n-1}d\mu_{i_1}(y)=\int_{0}^{1}z^{n-1}dz=\frac{1}{n}$$

$$Prob\left[R\left(s_{i_n},S_n\right)=1\right]=Prob\left[Y_{i_n}\geq Y_{i_k},\forall k=1,\cdots,n-1\right]=\prod_{k=1}^{n-1}Prob\left[Y_{i_n}\geq Y_{i_k}\right]=\int_{y^{min}}^{y^{max}}\prod_{k=1}^{n-1}Prob\left[y\geq Y_{i_k}\right]d\mu_{i_n}(y)$$

$$=\int_{y^{min}}^{y^{max}}\prod_{k=1}^{n-1}\mu_{i_k}(y)d\mu_{i_n}(y)<\int_{y^{min}}^{y^{max}}\left[\mu_{i_n}(y)\right]^{n-1}d\mu_{i_n}(y)=\int_{0}^{1}z^{n-1}dz=\frac{1}{n}$$

The two strict inequalities in the above two derivations come from the assumption of the first order stochastic dominance among independent student score distributions. □

Proof of Lemma 2.2

For any student pair s_i and $s_{i'}$, $i < i'$ implies that $\mu_i(\cdot)$ first order stochastically dominates (FOSD) $\mu_{i'}(\cdot)$ by Definition 2.1. By the definition of FOSD, $\forall y \in \left[y^{min}, y^{max}\right]$, $\mu_i(y) \le \mu_{i'}(y)$, thus we must have $y_i^{inf} \ge y_{i'}^{inf}$ and $y_i^{sup} \ge y_{i'}^{sup}$. Following the same analysis, $i \le i'' < i''' \le i'$ implies $y_i^{inf} \ge y_{i''}^{inf} \ge y_{i'''}^{inf} \ge y_{i'}^{inf}$ and $y_i^{sup} \ge y_{i''}^{sup} \ge y_{i'''}^{sup} \ge y_{i'}^{sup}$. Since s_i and $s_{i'}$ are competing and $i < i'$, we have $y_i^{inf} < y_{i'}^{sup}$. By combining these two set of inequalities, we have:

$$y_i^{sup} \ge y_{i''}^{sup} \ge y_{i'''}^{sup} \ge y_{i'}^{sup} > y_i^{inf} \ge y_{i''}^{inf} \ge y_{i'''}^{inf} \ge y_{i'}^{inf} .$$

Thus for any two students $s_{i''}$ and $s_{i'''}$ with $i \le i'' < i''' \le i'$, we have $y_{i''}^{inf} < y_{i'''}^{sup}$, which means that they are competing with each other. □

Proof of Proposition 3.1

Since SD is a special case of the Top Trading cycles (TTC) mechanism, all the results regarding the 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 are essentially the same as the classical BOS and SD mechanisms, with school priorities being the 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 the stable matching outcome in NE when the 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 fair 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 assumption, BOS implements the unique fair matching outcome in NE.

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

Proof of Proposition 3.2

We first prove that the pre-SD mechanism is strategy-proof. As a well-known result, the standard (post-) SD mechanism is strategy-proof, i.e., the truth-telling strategy is the dominant strategy for all the students (Abdulkadiroglu and Sonmez, 2003). That is, for any realization of joint student score distribution $\boldsymbol{y} = \left\{y_i : i = 1, \cdots, N\right\}$, for any student s_i, for any strategy profile (σ_i, σ_{-i}), we must have $u_i(\sigma_i^T, \sigma_{-i} \mid \boldsymbol{y}) \geq u_i(\sigma_i, \sigma_{-i} \mid \boldsymbol{y})$, where σ_i^T is student s_i's truth telling strategy, σ_i is student s_i's strategy, and σ_{-i} is the vector for all other students' strategies. Denote the joint score distribution function by $\mu(\cdot)$, thus we can have,

$$EU_i(\sigma_i^T, \sigma_{-i}) \equiv \int u_i(\sigma_i^T, \sigma_{-i} \mid \boldsymbol{y}) d\mu(\boldsymbol{y}) \geq \int u_i(\sigma_i, \sigma_{-i} \mid \boldsymbol{y}) d\mu(\boldsymbol{y}) \equiv EU_i(\sigma_i, \sigma_{-i}), \quad \forall \sigma_i, \sigma_{-i}.$$

Note that $EU_i(\sigma_i, \sigma_{-i})$ is exactly student s_i's expected utility under the pre-SD mechanism when the strategy profile is (σ_i, σ_{-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 playing in the post-SD mechanism, the matching outcome must be the same, which is the unique ex-post fair matching outcome. □

Proof of Proposition 3.3

The completely ex-ante fair matching outcome implies that student s_i is matched to school c_i, for any $i = 1, \ldots, N$. We want to prove that if the completely ex-ante fair matching outcome is implemented in NE under the pre-BOS mechanism, every student except student s_N (the least able student) puts his ex-ante fair matching school as his first choice.

Suppose instead that in the equilibrium where complete ex-ante fairness is achieved, there is one student s_i with $i < N$, who does not list his ex-ante fair matched school as his top choice. However, student s_i must be matched with his unique ex-ante fair matched school, c_i, in the equilibrium we assumed. This implies that student s_i is admitted by school c_i through his non-top choice, which in turn implies that there must be an empty slot at school c_i after the first round admissions under the pre-BOS mechanism.

Since $i < N$, there must exist some student $s_{i'}$ with $i < i' \leq N$, who prefers c_i to his ex-ante fair matched school $c_{i'}$. Note also that in equilibrium student $s_{i'}$ must be admitted by his ex-ante fair matched school $c_{i'}$. However, suppose now student $s_{i'}$ submits c_i as his top choice, then he will surely be admitted by c_i. Thus we find a profitable deviation for student $s_{i'}$, which invalidates the assumed equilibrium. □

Proof of Corollary 3.1 (of Proposition 3.3)

Consider a student s_i with $i < N$ who is playing a mixed-strategy under NE. Due to the uniqueness of completely ex-ante fair matching outcome, he must be admitted by school c_i for sure. Suppose that in the support set for his mixed 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 completely ex-ante fair, there should be no other students who list school c_i as their first choices in their mixed strategies. Otherwise student s_i would not be admitted by school c_i for sure. Now consider student s_N. In the equilibrium we assumed, s_N must be matched with school c_N. However, if he deviates from his equilibrium strategy to a strategy where he lists school c_i as his first choice, he will have positive probability of being admitted by c_i. Since c_N is the last-ranked school, he must be better off through such a deviation. This again invalidates the assumed equilibrium. □

Proof of Proposition 3.4

Sufficiency. Suppose that $\forall k = 1, \cdots, N$, $\forall i < k$ such that s_i and s_k are a pair of competing students, we have $Prob\left[Y_k > Y_i\right] \cdot u_k(c_i) \le u_k(c_k)$. Obviously, $\forall k = 1, \cdots, N$, $\forall i < k$ such that s_i and s_k are not a pair of competing students, we have $Prob\left[Y_k > Y_i\right] \cdot u_k(c_i) = 0 \cdot u_k(c_i) = 0 \le u_k(c_k)$. In the case $i > k$, we also have $Prob\left[Y_k > Y_i\right] \cdot u_k(c_i) \le u_k(c_i) < u_k(c_k)$. Combining all these three cases, the following condition holds for any $k = 1, \cdots, N$:

$$\forall i \neq k, \quad Prob\left[Y_k > Y_i\right] \cdot u_k(c_i) \le u_k(c_k), \text{ or } \underset{i \neq k}{Max}\, Prob\left[Y_k > Y_i\right] \cdot u_k(c_i) \le u_k(c_k).$$

We want to prove that if the above condition holds, the strategy profile where all the students list their ex-ante fair matched schools as their choices forms a Nash equilibrium.

Consider student s_k. Given that all other students list their ex-ante fair matched schools as their choices, if s_k deviates from that strategy, he will either be admitted by the school he lists as his (only) choice, or not be admitted at all ending up with a payoff of zero. Thus through deviation the highest payoff he can have is $\underset{i \neq k}{Max}\, Prob\left[Y_k > Y_i\right] \cdot u_k(c_i)$. If s_k sticks to strategy of listing his ex-ante fair matched school, his payoff is $u_k(c_k)$. By the condition above, s_k has no incentive to deviate at all.

Necessity. If the constrained pre-BOS/SD mechanism (where each student can only apply for one school) implements in its pure-strategy Nash equilibrium the completely ex-ante fair matching outcome, it can be easily shown that in equilibrium each student (including s_N) must list his ex-ante fair matched school as his (only) choice. In order to guarantee that no student has incentive to list any school rather than his ex-ante fair matched school, the following condition must hold:

$$\forall k = 1, \cdots, N, \forall i \neq k, \quad Prob\left[Y_k > Y_i\right] \cdot u_k(c_i) \le u_k(c_k).$$

In particular, the condition above must hold in the case that $i < k$ and s_i and s_k are a pair of competing students. □

Proof of Theorem 4.1

Denote the score rank cumulative distribution function (CDF) of student s_i by $\gamma_i(\cdot)$, we have $\gamma_i(k) \equiv Prob\left[R_i \le k\right],\ \ \forall k = 1,\cdots,N$. We first prove the following lemma.

Lemma A.1: If students' score distribution functions satisfy first-order stochastic dominance relationship, then their score rank distribution functions satisfy the (reversed) first-order stochastic dominance relationship. That is, $\forall i,i' \in \left\{1,...,N\right\}, i < i'$, ***if (i)*** $\forall y \in \left[y^{min}, y^{max}\right]$, $\mu_i(y) \le \mu_{i'}(y)$ ***and (ii)*** $\exists y' \in \left[y^{min}, y^{max}\right]$ ***such that*** $\mu_i(y') < \mu_{i'}(y')$, ***then (iii)*** $\forall k \in \left\{1,\cdots,N\right\}$, $\gamma_i(k) \ge \gamma_{i'}(k)$ ***and (iv)*** $\exists k' \in \left\{1,\cdots,N\right\}$ ***such that*** $\gamma_i(k') > \gamma_{i'}(k')$.

Proof:

To simplify our proof, here we further assume that for all students, the score distribution function, $\mu_i(\cdot)$, is non-atomic and differentiable within its support $supp(\mu_i) \subseteq \left[y_i^{inf}, y_i^{sup}\right]$. Let $\pi_i(\cdot)$ be its probability density function. That is, $\pi_i(y) > 0$ for any $y \in supp(\mu_i)$. First notice that:

$$Prob\left[R_i \le k\right] = \int_{y^{min}}^{y^{max}} \Gamma_i(k,y)\pi_i(y)dy,$$

where $\Gamma_i(k,y)$ is the probability that *at most* $k-1$ students in $S_{-i} \equiv \left\{s_j : j = 1,...,N, j \ne i\right\}$ have scores higher than y. Note that in the above equality $\pi_i(y) = 0$ for any $y \notin supp(\mu_i)$.

Consider two students s_i and $s_{i'}$, $i < i'$. We want to show the following results step by step.

Step 1. $\Gamma_i(k,y) \ge \Gamma_{i'}(k,y),\ \ \forall k \in \left\{1,\cdots,N\right\},\ \ \forall y \in \left[y^{min}, y^{max}\right]$.

We first introduce some notations. For $k \in \left\{1,\cdots,N\right\}$, let $E(k-1,-(i,i'),y)$ denote the event that at most $k-1$ students in $S_{-(i,i')} \equiv \left\{s_j : j = 1,...,N, j \ne i, j \ne i'\right\}$ have scores higher than y, and $E(k-2,-(i,i'),y)$ the event that at most $k-2$ students in $S_{-(i,i')}$ have scores higher than y. Obviously, we have $Prob\left[E(N-1,-(i,i'),y)\right] = Prob\left[E(N-2,-(i,i'),y)\right] = 1$. For convenience, assume $E(-1,-(i,i'),y)=\varnothing$ such that $Prob\left[E(-1,-(i,i'),y)\right]=0$.

Let $E(i,y)$ denote the event that student s_i has a score higher than y, and $E(i',y)$ the event that student $s_{i'}$ has a score higher than y. Let $E'(i,y)$ denote the event that student s_i has a score not higher than y, and $E'(i',y)$ the event that student $s_{i'}$ has a score not higher than y.

Now we can express $\Gamma_i(k,y)$ as the following:

$$\begin{aligned}\Gamma_i(k,y) &= Prob\left[E(k-1,-(i,i'),y)\right]\cdot Prob\left[E'(i',y)\right]+Prob\left[E(k-2,-(i,i'),y)\right]\cdot Prob\left[E(i',y)\right] \\ &= Prob\left[E(k-1,-(i,i'),y)\right]\cdot \mu_{i'}(y)+Prob\left[E(k-2,-(i,i'),y)\right]\cdot(1-\mu_{i'}(y))\end{aligned}.$$

It is easy to check that the above equality also holds if $k=1$ by our assumption.

Similarly, we have

$$\Gamma_{i'}(k,y)=Prob\left[E(k-1,-(i,i'),y)\right]\cdot \mu_i(y)+Prob\left[E(k-2,-(i,i'),y)\right]\cdot(1-\mu_i(y)).$$

Based on these two expressions, we can have

$$\Gamma_i(k,y)-\Gamma_{i'}(k,y)=\left[Prob\left[E(k-1,-(i,i'),y)\right]-Prob\left[E(k-2,-(i,i'),y)\right]\right]\cdot(\mu_{i'}(y)-\mu_i(y)).$$

$E(k-2,-(i,i'),y)\subseteq E(k-1,-(i,i'),y)$ implies $Prob\left[E(k-1,-(i,i'),y)\right]\ge Prob\left[E(k-2,-(i,i'),y)\right]$, $\forall k\in\{1,\cdots,N\}$, $\forall y\in\left[y^{min},y^{max}\right]$. By the FOSD assumption, $\mu_{i'}(y)\ge\mu_i(y)$ $\forall y\in\left[y^{min},y^{max}\right]$. Therefore, we have $\Gamma_i(k,y)\ge\Gamma_{i'}(k,y)$, $\forall k\in\{1,\cdots,N\}$, $\forall y\in\left[y^{min},y^{max}\right]$.

Step 2. $\boldsymbol{\int_{y^{min}}^{y^{max}}\Gamma_i(k,y)\pi_i(y)dy\ge\int_{y^{min}}^{y^{max}}\Gamma_i(k,y)\pi_{i'}(y)dy,\ \forall k\in\{1,\cdots,N\},\ \forall y\in\left[y^{min},y^{max}\right].}$

First, note that by definition $\Gamma_i(k,y)$ is a non-decreasing function of y. That is, $\frac{\partial\Gamma_i(k,y)}{\partial y}\ge 0$, $\forall y\in\left[y^{min},y^{max}\right]$.

Second, through integration by parts, we have

$$\int_{y^{min}}^{y^{max}} \Gamma_i(k,y)\pi_i(y)dy = \int_{y^{min}}^{y^{max}} \Gamma_i(k,y)d\mu_i(y) = \Gamma_i(k,y)\mu_i(y)\Big|_{y=y_i^{inf}}^{y=y_i^{sup}} - \int_{y_i^{inf}}^{y_i^{sup}} \frac{\partial\Gamma_i(k,y)}{\partial y}\mu_i(y)dy$$
$$= \Gamma_i(k,y_i^{sup}) - \int_{y_i^{inf}}^{y_i^{sup}} \frac{\partial\Gamma_i(k,y)}{\partial y}\mu_i(y)dy \qquad (1)$$

Similarly, we have

$$\int_{y^{min}}^{y^{max}} \Gamma_i(k,y)\pi_{i'}(y)dy = \Gamma_i(k,y_{i'}^{sup}) - \int_{y_{i'}^{inf}}^{y_{i'}^{sup}} \frac{\partial\Gamma_i(k,y)}{\partial y}\mu_{i'}(y)dy \qquad (2)$$

Since $\mu_{i'}(y) \geq \mu_i(y) \quad \forall y \in \left[y^{min}, y^{max}\right]$, we have $y_i^{sup} \geq y_{i'}^{sup}$, which implies that the first term of equation (1) is no less than the first term of equation (2), since $\Gamma_i(k,y)$ is non-decreasing in y. Since $\frac{\partial\Gamma_i(k,y)}{\partial y} \geq 0$ and $\mu_{i'}(y) \geq \mu_i(y) \quad \forall y \in \left[y^{min}, y^{max}\right]$, the second term (the integration part) of equation (1) is no larger than the second term of equation (2). Combining two observations together, we have the result:

$$\int_{y^{min}}^{y^{max}} \Gamma_i(k,y)\pi_i(y)dy \geq \int_{y^{min}}^{y^{max}} \Gamma_i(k,y)\pi_{i'}(y)dy, \quad \forall k \in \{1,\cdots,N\}, \quad \forall y \in \left[y^{min}, y^{max}\right].$$

Step 3. $\boldsymbol{\forall k \in \{1,\cdots,N\}, \quad \gamma_i(k) \geq \gamma_{i'}(k)}$**.**

$$\gamma_i(k) \equiv Prob\left[R_i \leq k\right] = \int_{y^{min}}^{y^{max}} \Gamma_i(k,y)\pi_i(y)dy \geq \int_{y^{min}}^{y^{max}} \Gamma_i(k,y)\pi_{i'}(y)dy \geq \int_{y^{min}}^{y^{max}} \Gamma_{i'}(k,y)\pi_{i'}(y)dy = Prob\left[R_{i'} \leq k\right] \equiv \gamma_{i'}(k).$$

The first inequality comes from Step 2 and the second inequality comes from Step 1.

Step 4. $\boldsymbol{\exists k' \in \{1,\cdots,N\}}$ ***such that*** $\boldsymbol{\gamma_i(k') > \gamma_{i'}(k')}$**.**

We consider two cases.

Case 1. Students s_i and $s_{i'}$ have no competing relationship.

In this case it is obvious that there exists some $k' \in \{1,\cdots,N\}$ such that $\gamma_i(k') > 0 = \gamma_{i'}(k')$.

Case 2. Students s_i and $s_{i'}$ have competing relationship.

According to Lemma 2.2, there must be some i'' and i''', satisfying $1 \leq i'' \leq i < i' \leq i''' \leq N$, such that all the students indexed from i'' to i''' have competing relationship with each other but all the other students (if exist) do not have competing relationship with all of them. Denote such a student set indexed from i'' to i''' by S', where $S' \equiv \left\{ s_j : i'' \leq j \leq i''' \right\}$. It is obvious that for any student $s_j \in S'$, their rankings in the whole set of students S only depend on their rankings within S'. We denote the number of students in S' by N', where $N' \equiv \#(S')$. Let $R'_j \equiv R(s_j, S')$ and $\gamma'_j(k) \equiv Prob\left[R'_j \leq k \right]$, for any $s_j \in S'$, and for any $k \in \left\{1, \cdots, N'\right\}$.

In Case 2, it suffices for us to show that for any two students s_i and $s_{i'}$ in S' with $i < i'$, there exists some $k' \in \left\{1, \cdots, N'\right\}$ such that $\gamma_i{}'(k') > \gamma'_{i'}(k')$. Note that this is almost the same problem as the original one, except that now all the students have competing relationship with each other. With a bit of abuse of notations, we keep the same notations as in the original problem if not specified.

By following our previous proof process, Note that it suffices to prove the following statement: $\exists y \in (y_{i'}^{inf}, y_{i'}^{sup})$, $\exists k \in \left\{1, \cdots, N'-1\right\}$, such that $\Gamma_i(k, y) > \Gamma_{i'}(k, y)$. This will guarantee that the second inequality in Step 3 would be strict.

Since $\Gamma_i(k, y) - \Gamma_{i'}(k, y) = \left[Prob\left[E(k-1, -(i,i'), y) \right] - Prob\left[E(k-2, -(i,i'), y) \right] \right] \cdot (\mu_{i'}(y) - \mu_i(y))$, we would like to show that for some y such that $y \in (y_{i'}^{inf}, y_{i'}^{sup})$ *and* $\mu_{i'}(y) > \mu_i(y)$, we have:

$$Prob\left[E(k-1, -(i,i'), y) \right] - Prob\left[E(k-2, -(i,i'), y) \right] > 0 \text{, for some } k, \ k \in \left\{1, \cdots, N'-1\right\}.$$

Let $E''(k-1, -(i,i'), y)$ denote the event that exactly $k-1$ students in $S_{-(i,i')}$ have scores higher than y, thus we have $Prob\left[E(k-1, -(i,i'), y) \right] - Prob\left[E(k-2, -(i,i'), y) \right] = Prob\left[E''(k-1, -(i,i'), y) \right]$.

Note that if for some $y \in (y_{i'}^{inf}, y_{i'}^{sup})$, $\mu_{i'}(y) > \mu_i(y)$, then there must exist some y' such that $y^{min} \leq y_{i'}^{inf} \leq y_i^{inf} < y' < y_{i'}^{sup} \leq y_i^{sup} \leq y^{max}$, and $\mu_{i'}(y') > \mu_i(y')$. This is because $\mu_{i'}(\cdot)$ and $\mu_i(\cdot)$ are differentiable, and $\pi_{i'}(\cdot), \pi_i(\cdot) > 0$ within their respective supports. So it suffices to prove the following statement:

Statement (A1): for *any* y such that $y_i^{inf} < y < y_{i'}^{sup}$, we have:

$$Prob\left[E''(k-1,-(i,i'),y)\right]>0 \text{ for some } k,\ k\in\{1,\cdots,N'-1\}.$$

We prove the above statement by considering three subcases 2.1-2.3.

Case 2.1. $N'=2$. That is, $S'=\{s_i,s_{i'}\}$. In this case we have $Prob\left[E''(0,-(i,i'),y)\right]=1>0$. So for any $y\in(y_i^{inf},y_{i'}^{sup})$, Statement (A1) holds for $k=1$.

Case 2.2. $N'>2$ and $i=i''$, that is, student s_i is the student with the highest ability in S'. We would like to show that $\forall y\in(y_i^{inf},y_{i'}^{sup})$, $Prob\left[E''(0,-(i,i'),y)\right]>0$ must hold; In other words, Statement (A1) holds for $k=1$ and for any $y\in(y_i^{inf},y_{i'}^{sup})$.

To see this, denote all the students in S' except s_i and $s_{i'}$ by a set $S'_{-(i,i')}\equiv\{s_{i^*}:i^*\neq i,i^*\neq i',i''\leq i^*\leq i'''\}$. By FOSD and the definition of competing relationship (Definition 2.2'), we have $y_{i^*}^{inf}\leq y_i^{inf}<y_{i^*}^{sup}\leq y_i^{sup}$, for any $s_{i^*}\in S'_{-(i,i')}$. Then for any y such that $y_i^{inf}<y<y_{i'}^{sup}$, we also have $y>y_{i^*}^{inf}$. This implies that $\forall y\in(y_i^{inf},y_{i'}^{sup})$, $\forall s_{i^*}\in S'_{-(i,i')}$, we have $Prob\left[Y_{i^*}\leq y\right]>0$. For example, each student s_{i^*} has a score y_{i^*} (and indeed a score range) such that $y_{i^*}^{inf}<y_{i^*}<y$. Therefore, we have $Prob\left[E''(0,-(i,i'),y)\right]>0$.

Case 2.3. $N'>2$ and $i>i''$, that is, there is at least one student who has a higher ability than s_i in S'. Suppose there are m such students, where $1\leq m\leq N'-2$. Denote all those students by a set $S'_{<i}\equiv\{s_{\hat{i}}:\hat{i}<i,,i''\leq\hat{i}\leq i'''\}$. By FOSD and the definition of competing relationship (Definition 2.2'), we have $y_{i'}^{inf}\leq y_i^{inf}\leq y_{\hat{i}}^{inf}<y_{i'}^{sup}\leq y_i^{sup}\leq y_{\hat{i}}^{sup}$, for any $s_{\hat{i}}\in S'_{<i}$.

It can be shown that it is possible to have $Prob\left[E''(k-1,-(i,i'),y)\right]=0$ for any $k\in\{1,\cdots,m\}$ and *some* $y\in(y_i^{inf},y_{i'}^{sup})$: For $y\in(y_i^{inf},y_{\hat{i}}^{inf})$, for any student $s_{\hat{i}}\in S'_{<i}$, we have $Prob\left[Y_{\hat{i}}>y\right]=1$. So the number of students having score larger than y cannot be less than m .

However, for $k=m+1$ and for any $y\in(y_i^{inf},y_{i'}^{sup})$, $Prob\left[E''(k-1,-(i,i'),y)\right]>0$ must hold. To see this, note first that if every student $s_{\hat{i}}\in S'_{<i}$ has a score $y_{\hat{i}}\in(y,y_{i'}^{sup})$, which is possible, then

all the m students in $S'_{<i}$ have scores larger than y; To put it formally, $\forall y \in (y_i^{inf}, y_{i'}^{sup})$, $\forall s_{\hat{i}} \in S'_{<i}$, we have $Prob\left[Y_{\hat{i}} > y\right] > 0$. Second, based on the analysis in Case 2.2, for any student except $s_{i'}$ who has a lower ability than s_i, it is possible that none of them has a score larger than y; To put it formally, $\forall y \in (y_i^{inf}, y_{i'}^{sup})$, $\forall s_{\tilde{i}} \in S'_{-(i,i')} - S'_{<i}$, we have $Prob\left[Y_{\tilde{i}} \le y\right] > 0$. Combining the above two facts, we have $Prob\left[E''(m, -(i,i'), y)\right] > 0$. Thus Statement (A1) holds for $k = m+1 \in \{1, \cdots, N'-1\}$ and for any $y \in (y_i^{inf}, y_{i'}^{sup})$. □

After proving Lemma A.1, it is straightforward to prove Theorem 4.1.

Proof of Theorem 4.1 (with Lemma A.1)

Let l_{ij} denote the probability of student s_i being ranked the j th by his realized score among all the students (i.e., $R_i = j$). By Propositions 3.1. and 3.2, all the three mechanisms (pre-SD, post-SD, and post-BOS) will always implement ex-post fairness. Therefore, $p_{ij} = l_{ij}$, $\forall i, j \in \{1, \cdots, N\}$. Recall that p_{ij} is the probability of student s_i being matched with school c_j.

By Lemma A.1, for any pair of students s_i and $s_{i'}$ with $i < i'$, we have $\forall k \in \{1, \cdots, N\}$, $\sum_{j=1}^{k} l_{ij} \equiv \gamma_i(k) \ge \gamma_{i'}(k) \equiv \sum_{j=1}^{k} l_{i'j}$ and $\exists k' \in \{1, \cdots, N\}$ such that $\sum_{j=1}^{k'} l_{ij} \equiv \gamma_i(k') > \gamma_{i'}(k') \equiv \sum_{j=1}^{k'} l_{i'j}$. Since $\forall i, j \in \{1, \cdots, N\}$, $p_{ij} = l_{ij}$, we have the following result: for any pair of students s_i and $s_{i'}$ with $i < i'$, (i) $\forall k \in \{1, \cdots, N\}$, $\sum_{j=1}^{k} p_{ij} \ge \sum_{j=1}^{k} p_{i'j}$ and (ii) $\exists k' \in \{1, \cdots, N\}$ such that $\sum_{j=1}^{k'} p_{ij} > \sum_{j=1}^{k'} p_{i'j}$. By Definition 2.5', this matching outcome is stochastically ex-ante fair. □

Proof of Proposition 4.1

We prove by contraposition. That is, under the n-degree ($n > 1$) competing relationship, pre-BOS will *not* implement stochastic ex-ante fairness in its weakly-dominant pure-strategy NE *if* $N > \dfrac{n(n+1)}{2}$. Our proof will focus on student's first choices. We will find conditions under which students' first choices in a weakly dominant pure-strategy NE will surely lead to the violation of stochastic ex-ante fairness. Let $j_c(s_i)$ be the index of the school that student s_i lists as his first choice, where $j_c(s_i) \in \{1,\cdots,N\}$, $\forall i \in \{1,\cdots,N\}$.

We first prove the following 4 lemmas.

Lemma A.2: In a weakly dominant pure-strategy NE where stochastic ex-ante fairness is achieved, students' first choices must be monotonic. Specifically, (i) $j_c(s_1)=1$, (ii) $0 \le j_c(s_{i+1}) - j_c(s_i) \le 1$, $\forall i \in \{1,\cdots,N\}$.

Proof:

(i) Suppose $j_c(s_1) > 1$. Then in a weakly dominant pure-strategy NE, there must be another student s_i with $i > 1$ such that $j_c(s_i) = 1$, and $p_{i1} > 0$. $j_c(s_i) = 1$ and $j_c(s_1) > 1$ imply $p_{11} = 0$ by the matching rules of the pre-BOS mechanism. Since $p_{i1} > p_{11}$ and $i > 1$, stochastic ex-ante fairness is violated.

(ii) For $i = 1,\cdots,N$, consider the very first i violating the condition $0 \le j_c(s_{i+1}) - j_c(s_i) \le 1$.

Suppose that the violation is $j_c(s_{i+1}) - j_c(s_i) < 0$. Obviously $i \ge 2$. Since for all $i' < i$, we have $0 \le j_c(s_{i'+1}) - j_c(s_{i'}) \le 1$ and $j_c(s_1) = 1$, for any school c_j with $j < j_c(s_i)$, there must be a student $s_{i'}$ such that $j_c(s_{i'}) = j$. $j_c(s_{i'}) = j$ and $j_c(s_i) > j$ imply $p_{ij} = 0$. Therefore, we have $\sum_{j<j_c(s_i)} p_{ij} = 0$. It is also known that in a weakly dominant pure-strategy Nash equilibrium, $p_{i+1, j_c(s_{i+1})} > 0$. However, $j_c(s_{i+1}) - j_c(s_i) < 0$ and $p_{i+1, j_c(s_{i+1})} > 0$ imply $\sum_{j<j_c(s_i)} p_{i+1,j} > 0 = \sum_{j<j_c(s_i)} p_{ij}$, violating stochastic ex-ante fairness.

Suppose instead that the violation is $j_c(s_{i+1}) - j_c(s_i) > 1$. Case (1): $i+1 < N$. Then in a weakly dominant pure-strategy equilibrium there is a student $s_{i'}$ with $i' > i+1$ such that $j_c(s_{i'}) = j_c(s_i) + 1$ and $p_{i', j_c(s_i)+1} > 0$. Thus $\sum_{j \le j_c(s_i)+1} p_{i'j} > 0$. However, $\sum_{j \le j_c(s_i)+1} p_{i+1,j} = 0$. This results in the violation of stochastic ex-ante fairness. Case (2): $i+1 = N$. Then student s_{i+1} will be better off by listing his first choice as $j_c(s_i)+1$ instead of $j_c(s_{i+1})$.

Lemma A.3. If the degree of competition is n ($\mathbf{n} > \mathbf{1}$), then there are at most n consecutively

indexed students whose first choice is the same school in the weakly dominant NE, if this school is not c_N ***or*** c_{N-1}.

Proof:

Suppose there are $n+1$ students $s_i, s_{i+1}, \cdots, s_{i+n}$ who choose school c_j with $1 \le j \le N-2$ as their first choice. Obviously $p_{i+n,j} = 0$. So in a weakly dominant NE, s_{i+n} should choose another school $c_{j'}$ with $j < j' \le N-1$, as his first choice. By Lemma A.2, those schools in $\{c_{j'} : j < j' \le N-1\}$ are the only schools that have not been chosen as the first choices of the students who have higher abilities than s_{i+n}. s_{i+n} will (weakly) benefit from this change.

Lemma A.4. For any $m < n$ ***consecutively indexed students whose first choice is the same school (not*** c_N ***or*** c_{N-1}***), if there exists some student indexed after them listing a different school as his first choice, there exist at least m+1 such students who are consecutively indexed and list that school as their first choice, to make stochastic ex-ante fairness possible.***

Proof:

Consider m consecutively indexed students whose first choice is the same school (not c_N or c_{N-1}). Denote the set of these students by $\{s_i, s_{i+1}, \cdots, s_{i+m-1}\}$, and their first choice by $c_{j'}$ where $j' < N-1$. By Lemma 2.1, for student s_{i+m-1}, we have $p_{i+m-1,j'} < \frac{1}{m}$. By Lemma A.2, if stochastic ex-ante fairness is achieved, we have $\sum_{j \le j'} p_{i+m-1,j} = p_{i+m-1,j'} < \frac{1}{m}$.

Consider the students indexed just after those m students. Suppose there are m' consecutively indexed students who have the same first choice, where $m' \ge 1$. Denote the set by $\{s_{i+m}, s_{i+m+1}, \cdots, s_{i+m+m'-1}\}$. By Lemma A.2, if stochastic ex-ante fairness is achieved, their first choice must be $c_{j'+1}$. Consider student s_{i+m}. By Lemma 2.1, we have $p_{i+m,j'+1} > \frac{1}{m'}$ and $\sum_{j \le j'+1} p_{i+m,j} = p_{i+m,j'+1} > \frac{1}{m'}$. Note that since $p_{i+m-1,j'+1} = 0$, then $\sum_{j \le j'+1} p_{i+m-1,j} = \sum_{j \le j'} p_{i+m-1,j} < \frac{1}{m}$. If $m' \le m$, then $\sum_{j \le j'+1} p_{i+m,j} > \frac{1}{m'} \ge \frac{1}{m} > \sum_{j \le j'+1} p_{i+m-1,j}$, in which case stochastic ex-ante fairness is violated. Therefore, in order to guarantee stochastic ex-ante fairness, we must have $m' \geq m+1$.

Lemma A.5. In a weakly dominant pure-strategy NE achieving stochastic ex-ante fairness, there cannot be 3 consecutively indexed students who list different schools from each other as their first choices.

Proof:

Suppose that there are 3 students s_i, s_{i+1}, s_{i+2} who list different schools as their first choices, say c_j, c_{j+1}, c_{j+2} (by Lemma A.2). Then by Lemma A.2, for any student $s_{i'}$ with $i' \neq i+1$, we must have $j_c(s_{i'}) \neq j+1$. Also, there will be no students who list c_{j+1} as their second choice in weakly dominant equilibrium, since c_{j+1} has been taken in the first round by student s_{i+1} (or, with others). In this case student s_{i+1} will have incentive to deviate by listing c_j as his first choice and c_{j+1} as his second choice. Obviously s_{i+1} can be better off by such a deviation.

Proof of Proposition 4.1 (with Lemmas A.2-A.5)

From Lemmas A.2-A.5, it is obvious that for competition degree $n > 1$, in stochastically ex-ante fair weakly dominant pure-strategy NE, the longest possible sequence of students with their first choices not violating stochastic ex-ante fairness would be as follows:

Student s_1 (one student) chooses c_1 as his first choice.

Students s_2 and s_3 (two students) choose c_2 as their first choice.

Students s_4, s_5 and s_6 (three students) choose c_3 as their first choice.

……

Students $s_{1+\sum_{k=1}^{n-1} k}, \cdots, s_{n+\sum_{k=1}^{n-1} k}$, (n students) choose c_n as their first choice.

These are in total $n + \sum_{k=1}^{n-1} k = \sum_{k=1}^{n} k = \frac{n(n+1)}{2} \equiv \bar{N}$ students. Note that for $n \geq 2$, we have $\bar{N} > n$, so c_n is a non-weakly dominated first choice for students. If there is one more (or further more) student, stochastic ex-ante fairness cannot be reached under the weakly dominant pure-strategy NE.

It is worth mentioning that if there are less students, we can always find a sequence of students with their first choices not violating the stochastic ex-ante fairness. To see this, denote the above longest list of student by $\bar{S}$, with a length of $\bar{N}$, and a corresponding first choice list as $\bar{C}_F$. Consider a student list S' with a length of $\bar{N}-m$, where $1 \le m < \bar{N}$. We construct a first choice list (say, C'_F) to S' as below: (i) We delete the first m choices from $\bar{C}_F$. (ii) In the remaining list, for any school indexed by j, we replace them with some single index j'. (iii) The new index is consecutive from 1 to $n-j$, for some $1 \le j < n$. (For example, a first choice list $(1,2,2,3,3,3)$ is the longest possible list not violating stochastic ex-ante fairness, with 6 students, if $n=3$. If we have only 5 students, we can construct a new first choice list $(1,1,2,2,2)$ for them, which does not violate stochastic ex-ante fairness. □