China Economic Review 79 (2023) 101963

Contents lists available at ScienceDirect

China Economic Review

journal homepage: www.elsevier.com/locate/chieco

When do stable matching mechanisms fail? The role of standardized tests in Chinese college admissions

Ming Jiang [a], Jingchao Li [b,*]

[a] Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, China
[b] School of Business, East China University of Science and Technology, Shanghai, China

ARTICLE INFO

JEL classification:
C78
C92
D47
I28

Keywords:
School choice matching
Chinese college admissions
Market unraveling
Experiment

ABSTRACT

In this paper, we investigate matching problems where priorities and preferences are misaligned. In the case of centralized Chinese college admissions, students are matched based on their test scores in standardized tests, a noisy realization of their aptitudes due to measurement errors. We show that in this case any matching mechanism that is stable with respect to score is not stable with respect to aptitude. The resulting instability leads to colleges' incentive to participate in early admissions (zizhu zhaosheng), a form of market unraveling. However, a manipulable mechanism such as the Immediate Acceptance mechanism, combined with limited information about priorities, may succeed in mending this market failure. We then design and conduct a laboratory experiment where we compare the performance of two mechanisms (the Immediate Acceptance mechanism and the Deferred Acceptance mechanism), under two timing conditions of the submission of students' rank-ordered lists of colleges (before the exam and after the exam), using a two-stage matching market design with the possibility of early offers. In the experiment, a significant level of market unraveling occurred under mechanisms that are not stable with respect to aptitude, confirming theoretical predictions. We also find that the Immediate Acceptance mechanism under pre-exam submission condition significantly reduces such unraveling.

1. Introduction

College admissions mechanisms affect the career choices and labor market outcomes of many young people in China and around the world (Balinski and Sönmez, 1999; Gale and Shapley, 1962; Roth, 1985). They belong to a broader class of matching problems that involve pairing members of one group of agents with one or more members of another disjointed group of agents. Other examples include matching medical school graduates to hospitals (Roth, 1984, 1986), assigning students to public schools (; Abdulkadiroğlu, Pathak, and Roth, 2005; Abdulkadiroğlu and Sönmez, 2003), assigning students to on-campus housing and overseas trips (Abdulkadiroğlu and Sönmez, 1999; Chen and Sönmez, 2002; Featherstone, 2020), facilitating pairwise kidney exchanges (Roth, Sönmez, and Ünver, 2005), and matching cadets to army branches (Sönmez and Switzer, 2013). *Matching mechanisms* are algorithms used to accomplish assignments in these cases.

For such matching problems, stable and strategy-proof matching mechanisms such as the Deferred Acceptance mechanism[1] (Gale

* Corresponding author.
E-mail addresses: mjiang@sjtu.edu.cn (M. Jiang), jingchaoli@ecust.edu.cn (J. Li).

[1] Note however, the Deferred Acceptance mechanism is strategy-proof for one side of the matching market. It is not so for the other side of the market.

https://doi.org/10.1016/j.chieco.2023.101963
Received 28 September 2022; Received in revised form 16 January 2023; Accepted 13 March 2023
Available online 24 March 2023

and Shapley, 1962) are widely praised as a superior alternative to priority matching mechanisms, such as the Boston mechanism (also known as the Immediate Acceptance mechanism, or IA henceforth, in college admissions), which is manipulable and may thus lead to unstable outcomes (Abdulkadiroğlu, Pathak, Roth, and Sönmez, 2005; Chen, Jiang, and Kesten, 2020; Chen and Sönmez, 2006; Ergin and Sönmez, 2006). This is because if a matching mechanism is stable, it produces matching outcomes in which no pair of agents would both prefer to be matched to each other than to their current partners. If a mechanism is strategy-proof, then participants do not have incentives to lie about their true preferences. A significant issue with unstable matchings is market *unraveling* (Dargnies, Hakimov, and Kübler, 2019; Roth and Xing, 1994): the transaction date is pushed further and further ahead, often circumventing the centralized matching process. For example, in the case of matching new physicians to hospitals, when the job market unravels, hospitals sign employment contracts with medical school students sometimes years ahead of their actual graduation date. Market unraveling also manifests itself through early admissions in college admissions (Avery, Fairbanks, and Zeckhauser, 2009). Roth (1991) shows that if a mechanism frequently yields unstable matchings, even if such market is centralized, it still causes market unraveling, and is no better than the decentralized market it tries to replace. Such unraveling is believed to be inefficient (Fréchette, Roth, and Ünver, 2007), and reduces market mobility (Niederle and Roth, 2003). In addition to the cost of instability, Ergin and Sönmez (2006) and Chen and Sönmez (2006) also suggest that strategic manipulation in the Boston mechanism leads to efficiency loss both in theory and in the laboratory.

Stability, however, does not ensure that market unraveling would not happen. One instance from two-sided matching comes from Sönmez (1999), which states that stability does not prevent manipulation through pre-arranged match. Even for centralized college admissions, which is a one-sided matching market, market unraveling still occurs under stable mechanisms. As an anecdotal example, the government of Shanghai in 2008 switched its college admissions mechanism from a variant of the Immediate Acceptance mechanism called the “sequential mechanism” to a variant of the Deferred Acceptance mechanism, known in the matching literature as the “Chinese Parallel Mechanism,” which is more stable and less manipulable (Chen and Kesten, 2017).[2] Yet, after the policy change, the top universities in Shanghai increased the number of their early admission offers from 893 to 1353,[3] an almost 50% increase, which all happened months before the actual college entrance exam, circumventing the centralized matching process. While centralized Chinese college admissions are hailed as a very efficient process for matching students and colleges (Chen et al., 2020; Chen and Kesten, 2017, 2019), the decentralized early admissions put significant additional burdens on college administrators and student families alike, as each participating colleges need to administer their own tests and employ hundreds of experts to interview students individually, and students need to prepare for different colleges' tests separately. Of course, there are many potential reasons for a large increase in the number of early admissions, and it is not the focus of this study to provide an empirical evaluation of the effect of such policy change. In this paper, we propose one explanation for the unraveling that arises from the “side-effect” of stable matching mechanisms under the context of standardized test, and test such explanation experimentally, as it is difficult to collect data to test it empirically.

To explain why such unraveling happened, first we need to note that college admissions in China are organized as a centralized matching clearinghouse,[4] where a central authority administers an annual standardized test (gaokao), and places students into colleges based on their test scores, with the aim of the test being to measure students' aptitudes. This standardized test-based college admissions problem differs from the school choice problem (Abdulkadiroğlu and Sönmez, 2003) in that although colleges, like public schools, follow priorities based on test scores, colleges themselves actually have strong preferences over students that may be different from priorities. It is understandable that colleges prefer better students with higher academic abilities (aptitudes) as better students are more likely to improve college reputations and may bring more future donations. It also differs from decentralized college admissions systems like those in the United States (Gale and Shapley, 1962; Roth, 1985), because colleges can not express their preferences (with respect to aptitudes) over students once the centralized matching process starts, since the matching is carried out by a computer program and colleges are bound to honor students' ranks in their test scores, which determine their admission priorities. Colleges' preferences and students' priorities are perfectly aligned only if standardized tests yield scores that do not distort the relative standings of students' aptitudes. In practice, however, this is unlikely, because even if a test is unbiased, that is, the *expected* test scores always respect the relative standings of aptitudes, all tests have *measurement errors*, defined as the lack of “consistency with which the [test] results place students in the same relative position if the test is given repeatedly” (Bloom, Madaus, Hastings, et al., 1981). Therefore, if a test is taken only once, such as in centralized college admissions, the ranking of test scores may not reflect the ranking of aptitudes.

To see why measurement errors in tests can cause unstable matchings, which further lead to unraveling, let us look at an example. Suppose Students A and B are two top students in a college admission market with many students, and Colleges 1 and 2 are the two top colleges, and each has one available seat. Student A is better than B, and College 1 is better than 2. All students prefer better colleges, and colleges better students. Although Student A is better than Student B, she still has 30% of chance to score lower in the test than B, caused by measurement errors in the test. Under the Deferred Acceptance mechanism, since revealing one's true preference is the dominant strategy, both students list College 1 as their first choices, and College 2 s choices. Since priorities are given based on the test scores, therefore, there is a 30% chance that Student A goes to College 2 and Student B goes to College 1.[5] In this case, both Student A and College 1 prefer each other to what they are currently matched with. So College 1 has an incentive to circumvent the centralized

[2] We will henceforth simply refer to them as the Immediate Acceptance mechanism and the Deferred Acceptance mechanism respectively.

[3] http://edu.qq.com/a/20081208/000066.htm. Retrieved on 8/25/2014.

[4] See the online appendix to Chen and Kesten (2017) for a thorough review of the histories and evolutions of Chinese college admissions from a market design perspective.

[5] This, and the example in the next paragraph assume that no other student can score higher than both Students A and B under any circumstance.

matching, and admits Student A early; and Student A certainly has the incentive to accept such offer.

The above example may naturally lead to the following question: what could possibly be the mechanism that led to the significant increase of early admissions in Chinese gaokao? The answer may lie in one policy detail: instead of submitting their rank-ordered lists of colleges after knowing their test scores and rankings, students report their preferences before taking the exam, therefore the only thing they know is the relative standings of their academic abilities. Let us go back to the example above. Now when Student B submits his list of colleges without knowing his realized test score, he knows that he has a 70% chance to score lower than A. Since under the Immediate Acceptance mechanism, emphasis is put on the first choice, and the cost of not getting into one's first choice is very high, Student B may not want to take the risk and rank College 1 first, and therefore he will rank College 2 as his first choice. In this case, Student A can go to College 1, and Student B to College 2, with certainty. The matching is therefore stable for the top two students and colleges. So in short, the manipulability of this mechanism, combined with limited information of the priorities (test scores) of the matching, actually helps better students go to better colleges, regardless of their test performances. This in turn reduces market unraveling.

Previous literature has also shown, in different aspects, that the Immediate Acceptance mechanism could outperform the Deferred Acceptance mechanism in some aspects under certain conditions. In terms of market efficiency, the IA mechanism is shown to be ex ante more efficient under incomplete information than the DA mechanism both theoretically (Abdulkadiroğlu, Che, and Yasuda, 2011) and experimentally (Featherstone and Niederle, 2016). In terms of strategy, contrary to the Deferred Acceptance mechanism, the Immediate Acceptance mechanism is immune to manipulation by schools through misreporting enrollment capacity (Kesten, 2012). In addition to the Boston (IA) mechanism, other manipulable mechanisms can also take the advantage of their manipulability and perform better than truthful mechanisms. For example, manipulable mechanisms that encourage reporting indifferences can outperform strategy-proof Random Serial Dictatorship mechanism in terms of efficiency (Fragiadakis and Troyan, 2019). This paper also contributes theoretically to a strand of literature that looks at the side effects of strategy-proof and stable matchings (Abdulkadiroğlu et al., 2011; Abdulkadiroğlu, Che, and Yasuda, 2015; Featherstone and Niederle, 2016; Troyan, 2012).

In addition to the properties of mechanisms, the context of this study - centralized college and high school admissions in China - has also been explored in the literature, both experimentally (Chen and Kesten, 2019), and empirically (Chen et al., 2020; Ha, Kang, and Song, 2020; Wang and Zhou, 2020; Wu and Zhong, 2020). This study is also closely related to a strand of studies for Chinese college admissions that focus on the timing of rank-ordered list submissions. In their empirical study, Wu and Zhong (2014) look at students' academic performances after they were admitted to a college with different mechanisms across years and provinces to test the effect of submission timings. Lien, Zheng, and Zhong (2016) use laboratory experiments to show that the Boston mechanism combined with pre-exam submission has advantages in ex-ante fairness (i.e. stability w.r.t. aptitude) compared to the Serial Dictatorship mechanism, while Lien, Zheng, and Zhong (2017) provide a theoretical foundation for the experimental results, and show mixed results with the combination of Boston mechanism and pre-exam submission. Pan (2019) further expands by looking the effect of timing when information about students' academic aptitudes are uncertain and some students are overconfident. The main contribution of this paper compared to the existing literature is to study the consequences of instability (i.e. marketing unraveling) through a two-sided experimental design of the matching markets.

In this paper, motivated by the real world observation of unraveling of a college admissions matching market, we answer two questions: why the Deferred Acceptance mechanism can produce unstable matchings and cause market unraveling; and how the Immediate Acceptance mechanism can outperform the Deferred Acceptance mechanism in mending such market failure. To answer the question of *why*, we formally model the game of college admissions with standardized tests, and show that without the possibility of early admissions (a form of market unraveling), a stable matching mechanism that respects test scores produces unstable matching outcomes in terms of student aptitude. This directly predicts market unraveling, as most cases of market unraveling are caused by perceived instability in matching outcomes. We then design and conduct a laboratory experiment to answer the question of *how*. We combine the two mechanisms (the Immediate Acceptance mechanism and the Deferred Acceptance mechanism) with two timing conditions of the submission of students' rank-ordered lists of colleges. The environment is designed as a hybrid two-sided market, where there is a centralized, one-sided matching stage, and preceding it is a decentralized, two-sided stage where colleges can send early admission offers to students. The experimental results confirm the theoretical predictions, and show that the Immediate Acceptance mechanism combined with pre-exam submission performs best in terms of reducing market unraveling, yet it achieves that with little sacrifice on the stability of eventual matching outcomes.

The remainder of this paper is organized in the following way: Section 2 formally models the game of college admissions with standardized test as well as introducing the two mechanisms. Section 3 presents theoretical results. Section 4 describes the design of the experiment. Section 5 presents experimental results. Section 6 concludes.

2. The environment

2.1. The model

We outline a matching economy with a continuum of students to be assigned to a finite number of colleges.

The set of students is denoted by S, with any student s whose type is represented by his aptitude $a \in \mathbb{R}$. The total mass of students is normalized to 1, and the distribution of their aptitudes follows a density function $f(\cdot)$. Each student also has a test score t after taking the entrance test. A *standardized test* determines the relationship between the test score and the aptitude: $t = a + \eta$, where $\eta \in \mathbb{R}$ has a distribution with density function $g(\cdot)$, and a and η are independent. It captures the measurement error of the test, and is individually determined for each student. We assume that η has mean 0. That is, a test is unbiased.

The set of colleges is denoted by $C = \{c_1, \ldots, c_m\}$, $m \geq 2$. Each college c_i has capacity to enroll a mass of q_{c_i} students. Without loss of generality, we assume that the mass of total quota equals to the mass of students.

A matching μ is an allocation of college slots to students such that the mass of students assigned to any college does not exceed its quota. A matching mechanism accomplishes this matching using students' reported rank-ordered lists (ROL) of colleges, and students' *priorities*. In college admissions, students' priorities at every college during the admissions are determined by their test scores: s has a higher priority than s' if and only if $t > t'$. We also assume that all colleges have the same preferences over students, which are determined by students' aptitudes.

A student also has ranking in aptitude, r_a and ranking in test score, r_t, measured by the mass of students with higher aptitudes and test scores, respectively.

We also make the following assumptions: the distributions of aptitude and measurement error of the test (captured by $f(\cdot)$ and $g(\cdot)$) are common knowledge to all students and colleges; every student knows his ranking of test score after the exam; all students have the same preferences over colleges; students know their own values (and by extension, rankings) of aptitude, and there exists a (costly) common measure of aptitude, such that whoever applies this measure on a student will get the same (true) measure of the student's aptitude.

The assumption that students have the same preference over colleges is a simplification based on the nature of Chinese college admissions: good colleges are sought after by almost all students and students' preferences are highly correlated. The assumption that students know their own rankings of aptitude can be justified by the fact that students undergo many mock testings organized by their high schools and districts before the final gaokao, therefore they should have good ideas about where they stand within each school and district, and they are able to extrapolate that knowledge to the geographical units that gaokao is organized for.

Next, since both test score-based priorities and aptitude-based preferences coexist, we define two sets of stability criteria.

A matching μ is *stable with respect to test score* if there is no student-college pair (c, s) such that student s prefers college c to the college he is assigned to, and college c has students assigned to it who are ranked lower in test scores than student s.

A matching μ is *stable with respect to aptitude* if there is no student-college pair (c, s) such that student s prefers college c to the college he is assigned to, and college c has students assigned to it who have lower aptitudes than student s.

The definitions of stability w.r.t. test score and aptitude are closely related to those of ex-post and ex-ante fairness in one-sided college admissions matching market described in Lien et al. (2016).

2.2. The mechanisms

In this section we describe the algorithm of the two mechanisms we study in this paper.

The first mechanism we study is the Immediate Acceptance mechanism (IA). The procedure of this mechanism is listed below:

There is a priority ordering of students. In the case of centralized college admissions, it is determined by the ranks of students' test scores.

Step 1: The first choices of the students are considered. For each college, consider the students who have listed it as their first choice and assign seats of the college to these students one at a time following their priority order until either there is no seat left or there is no student left who has listed it as her first choice.

Step k ($k > 1$): For the students who have been rejected after step k-1, only the kth choices of them are considered. For each college with available seats, consider those students who have listed it as their kth choice and assign the remaining seats to these students one at a time following their priority order until either there is no seat left or there is no student left who has listed it as her kth choice.

The algorithm terminates when there is no rejected student. The IA mechanism is manipulable (Abdulkadiroğlu, Pathak, Roth, and Sönmez, 2005; Chen and Kesten, 2017; Ergin and Sönmez, 2006), and its manipulability has also been observed experimentally (Chen and Kesten, 2019; Chen and Sönmez, 2006) and empirically (Chen et al., 2020); particularly, it puts heavy emphasis on how students list their first choices, since admission at every step is final. Such manipulability therefore often yields matching outcomes that are unstable (Chen et al., 2020; Chen and Sönmez, 2006; Ha et al., 2020; Roth, 1991).

The second mechanism studied in this paper is the Deferred Acceptance mechanism (DA) (Gale and Shapley, 1962). This mechanism played a key role in school choice reforms in Boston and New York City (Abdulkadiroğlu et al., 2005b,a). The procedure of the DA mechanism is listed as the following algorithm.

For each college, a priority ordering of students is determined by the ranks of students' test scores.

Step 1: The first choices of the students are considered. For each college, consider the students who have listed it as their first choice and temporarily assign seats of the college to these students one at a time following their priority order until either there is no seat left or there is no student left who has listed it as her first choice.

Step k ($k > 1$): Each student who was rejected applies to the next college on her list. Each college then considers the students it has already temporarily accepted along with the new applicants, accept the ones with the highest priorities within their enrollment capacities among those students, and then rejects the rest.

The algorithm terminates when there is no rejected student. The DA mechanism is strategy-proof (Dubins and Freedman, 1981), and it also produces stable matchings that are most favorable to students (Roth, 1982).

3. Theoretical results

In this section, we present the theoretical predictions for the two mechanisms under the test-based college admissions environment. Since the stability and incentive compatibility of the Deferred Acceptance mechanism with deterministic priorities are well established

(Dubins and Freedman, 1981; Roth, 1982), we will start with the incentives for truth-telling under DA when students submit their rank-ordered list of colleges (ROL) before they take the exam. In this case, they only know their rankings in aptitude and the distribution of measurement error. It turns out that the strategy-proof property of DA still holds.

Proposition 1. (Roth, 1989). *It is a dominant strategy to reveal one's true preference under DA with pre-exam ROL submission.*

The next proposition is the extension of the standard stability result for the DA mechanism.

Proposition 2. The DA mechanism with pre-exam ROL submission yields stable matchings with respect to test score under the Bayesian Nash equilibrium in dominant strategies.

Proof. . First, it is straightforward to see that the set of stable matching (with respect to test score) is a singleton, because all students have identical priorities at every college, which are determined only by test scores. Next, by Proposition 1, we see that truth-telling is a dominant strategy under DA pre-exam, therefore truth-telling is the Bayesian Nash equilibrium in dominant strategies. Moreover, it is unique, because every other strategy is dominated by truth-telling. Finally, DA yields stable outcome (w.r.t. test score) when everyone is truth-telling.

Different from DA, the IA mechanism encourages strategic manipulation. Specifically, students manipulate their first choices based on their rankings in test scores. We define the following kind of strategic manipulation under the test-based college admissions, which has been observed in the laboratory (Chen and Kesten, 2019) as well as in the field (Chen et al., 2020).

Definition 1. (Rank bias). *A student s with rank r_t exhibits rank bias, if he ranks the least commonly preferred college as his first choice among all the colleges where the total quotas of all more commonly preferred colleges do not exceed his ranking in test score (r_t). That is, a student lists college c_i as his first choice if $\sum_{k=1}^{i-1} q_{c_k} < r_t$, $\sum_{k=1}^{i} q_{c_k} \geq r_t$.*

In other words, it means a student will rank a college whose rank in quality, quota considered, corresponds to the student's rank in test score. The definition of rank bias is closely related to the concept of district bias in the school choice literature (Chen and Sönmez, 2006). The following result shows that it is a strategy under IA with post-exam ROL submission. Appendix A.1 shows theoretically why rank bias is a Nash Equilibrium strategy under IA.

The consequence of the above results is that the relative standings of aptitude are totally ignored during the matching. For DA, it is ignored because it is always best to reveal one's true preference, and assignment process depends on priorities that are entirely decided by test scores. It is also ignored under IA, because as long as students know their test scores, the rankings of test scores are the only thing they need in order to strategize their first choices.

The following theorem then illustrates the consequences of such matching outcomes, when there are a continuum of students whose aptitudes are distributed following $f(\cdot)$, students and colleges have homogeneous preferences over each other, and test scores are noisy realizations of students' aptitudes with measurement error following $g(\cdot)$. It shows the conflicting nature of the two stability definitions.

Theorem 1. In the continuum economy of the test-based college admissions, with a continuum of students whose aptitudes are

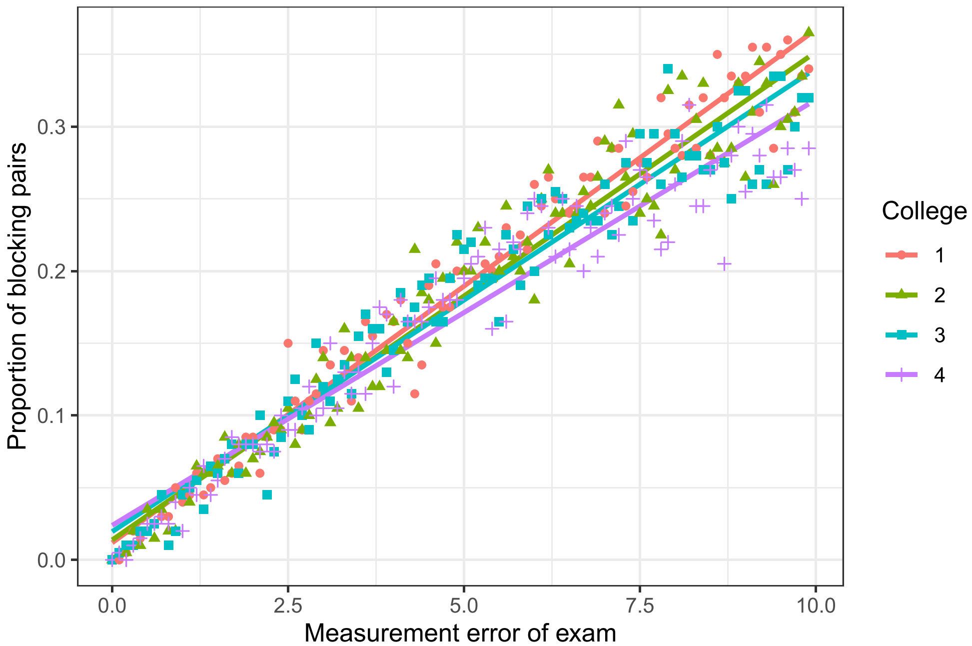

Fig. 1. Number of blocking pairs increases with measurement error.

distributed following the density function $f(\cdot)$, any matching outcome that is stable with respect to test score is not stable with respect to aptitude with probability 1. Specifically, the proportion of students each college prefers to be matched with rather than what it is matched with after the centralized admissions is non-zero.

Proof. . See Appendix A.2.

Theorem 1 provides a non-parametric look at the potential unstable outcomes of the matching. Intuitively, how large the measurement error of a test is, affects how unstable the matching outcomes are, with respect to aptitude. To give a concrete example, we conduct a simulation with a set of parameters. In this simulated example, there are 1000 students and five colleges, each with 200 seats. Students' aptitudes are uniformly distributed from 50 to 100, and the measurement error follows a normal distribution with mean 0. All students report their true preferences and are matched using the Deferred Acceptance mechanism. Fig. 1 reports the number of students with whom colleges would form blocking pairs, after varying the standard deviation of the distribution of measurement error (from 0 to 10). From this graph we can see that the proportion of blocking pairs (out of each college's capacity) increases as the standard deviation of measure error increases. Fig. A.2 in the appendix goes into what actually happened behind each dot in the previous graph by showing the distribution of aptitudes of students admitted by each college in a single simulation.

Since DA under truthful preference revelation, and IA with post-exam submission under the Nash Equilibrium strategy both yield the same matching outcomes that are stable with respect to test scores, we now have the following corollary.

Corollary 1. In the test-based college admissions, DA under post-ROL condition, and IA under post-exam condition yield Nash equilibrium matching outcomes that are not stable with respect to aptitude. Moreover, DA under pre-exam condition has a Bayesian Nash equilibrium in dominant strategies that yields an outcome unstable with respect to aptitude.

One of the most important consequences of matchings that are not stable w.r.t. aptitude (i.e. preference) is market unraveling. This phenomenon has been observed in many real world matching markets where matching mechanisms produce unstable outcomes. Examples include matching of couples in the National Resident Matching Program (Roth, 1984), sorority rush at American universities (Mongell and Roth, 1991), new physicians matching in Newcastle, Edinburgh, Birmingham, UK (Roth, 1991), and an experimental labor market (Dargnies et al., 2019). In the case of Chinese college admissions, early admission (zizhu zhaosheng) is another form of such unraveling, where colleges organize separate tests and interviews, and send students early offers to secure seats ahead of centralized admissions with entrance exam gaokao.[6] Similar two-stage market design has been used to model unraveling in the theoretical literature studying matching markets (Echenique and Pereyra, 2016). The following result shows why colleges have incentives to participate in early admissions, and how market can unravel as a result.

Proposition 3. (Unraveling). All but the last ranked colleges have the incentive to participate in early admissions when centralized matching is unstable with respect to aptitudes.

Proof. . See Appendix A.3.

While the instability with respect to aptitude under the DA mechanism is due to the fact that DA is strategy-proof, and allocation is entirely based on priorities, the reason for instability under the IA mechanism is different. Under the IA mechanism, students manipulate based on their priorities. When test scores are revealed after the exam, and ranks are known, aptitude plays no role in decision making. However, when scores are not known, and aptitude and the rank of it are the only clues students can follow to make their decisions, the fact that the IA mechanism is manipulable may make students reveal more information about their true aptitudes through submitted ROL. This has been shown to be the case in some matching environments in the literature, such as the concept of "fairness-revealing strategy" in Lien et al. (2016). The actual strategies students may employ under this mechanism-timing combination depend on their cardinal utilities as well as their risk preferences. Therefore, finding and generalizing the equilibrium strategies is challenging. In the following section, the matching market in the experiment design serves as an example, showing that in some cases, the IA mechanism with pre-exam submission indeed outperforms all other combinations in terms of stability w.r.t. aptitude.

4. Experimental design

The previous section outlines the theoretical stability results under different combinations of mechanisms and submission timings. The equilibrium under the IA mechanism with pre-exam submission, however, is difficult to derive for a general case. In this section, we design a specific, discrete market, where equilibrium strategies can be reasonably derived for the IA mechanism, and test the associated theoretical predictions using a laboratory experiment.

We implement a 2 (timing) by 2 (mechanism) factorial design to investigate the performance of the two mechanisms under two different timing conditions regarding when ROLs are submitted (pre-exam and post-exam). The four treatments are henceforth referred to as *IA-pre*, *IA-post*, *DA-pre*, and *DA-post*.

[6] There are other reasons that cause colleges to prefer early admissions. For example, Avery and Levin (2010) point out that colleges prefer early admissions because they provide students with opportunities to signal their interests. In this paper, we specifically focus on early admissions as a form of market unraveling.

4.1. The environment

The game environment consists of three students, {1,2,3}, and three colleges, $\{A,B,C\}$. Each college has exactly one slot, and each student can occupy at most one slot. Three students are ranked by their aptitudes, $1 \succ_c 2 \succ_c 3$, and three colleges their qualities, $A \succ_s B \succ_s C$. Students' and colleges' preferences over each other are described by utilities: {30,20,10}, which is the experimental points they earn given the students or colleges they are matched with. Matching payoffs for both students and colleges are summarized in Table 1.

In the theoretical derivation, we show that colleges have incentives to participate in early admissions if they expect matching outcomes to be unstable w.r.t. aptitude. In the experiment, colleges are allowed to make early offers, which captures the consequences of market instability. Therefore, the game in this experiment consists of four stages: Early Admission Stage, Exam Stage, ROL-submission Stage, and Centralized Matching Stage. Under the pre-exam timing condition, the game follows the following order: EarlyAdmission-ROL-Exam-Matching, while under the post-exam condition, the game follows EarlyAdmission-Exam-ROL-Matching.

Before students take the exam and submit their ROLs, colleges are given the opportunities to send binding early admission offers to students by paying a fee of 3 utility points. The fee represents the cost associated with the consequences of market unraveling[7]; as a result, colleges also get to know students' rankings of aptitudes (from interviewing, testing, reviewing their high school transcripts, etc.). The choice of 3 utility points is based on the consideration that it be close to the predicted incentives, such that while college players have the incentive to pay the cost to participate in early admissions, they would need some calculation to realize the fact. After paying the fee (i.e., organizing the early admission procedure), each college can send exactly one offer to any student. Students who receive the offers can choose to accept or decline the offers. Students and colleges that have reached the early admission agreements are removed from subsequent centralized matching market. The design of the early admissions stage is similar to Kagel and Roth (2000). It is a simplified environment of its real-world counterpart. The underlying process goes as following: the colleges pay the cost to organize an “early admission event”, where they get to reveal the aptitude measure of any students they are interested in, and then send offers to any students within the limitation of their admission quota. Since in the experimental environment, students' aptitude is determined by their type, and each college only has one slot, we omit the intermediate process.

During the Exam Stage, the computer simulates the exam by giving each student a test score. The measurement error of the test is characterized by the probability parameter p as well as the spread from the mean. Each student can perform normally with probability p, or either overperform or underperform each with probability $\frac{1-p}{2}$. Students know their types (aptitudes) and the distribution of test outcomes of all other students. The detailed testing outcome for the students are shown in Table 2. In this experiment we choose $p = 0.5$.

Either after or before taking the exam, depending on the timing condition, students who have not been admitted early by the colleges submit their ROLs of all three colleges, regardless whether the slot in some colleges are taken or not. If a college has reached an agreement and admitted a student, that college is removed from each student's ROL and does not occupy a position in students' ROLs. For example, Student 1 lists $A - B - C$ and College A has already admitted Student 2, then College A will simply be removed before the algorithm starts and College B becomes Student 1's first choice in the subsequent matching. This design is chosen to let college players think about, through repeated play, what would have happened if they choose to or not to participate in early admissions, since we allow them to observe the past ROL choices by the students. For student players, it also removes the influence of what other subjects have done in this round, so that they need to think for different contingencies.

One concern we may have, with respect to still allowing students to rank all three colleges even though one or more colleges have admitted students early, is that students' strategies may alter colleges' early admissions decisions. However, we will now show that as long as a student believes that a college, who should participate in early admissions, does not do so with any positive probability, his best response is to play the strategy as if no college participates in early admissions.

Under the post-exam condition, Student 3 is the only one who will make decisions after early admissions. Suppose he believes College A will not participate with probability $p > 0$ and College B with probability $q > 0$. Then he believes he plays the matching game with three colleges with probability pq, with Colleges B and C with probability $(1 - p)q$, and with Colleges A and C with probability $(1 - q)p$. Under DA, there are multiple best response strategies for the second and third scenario toward which Student 3 is indifferent. However, the only common best response for all three scenarios is the strategy of truth-telling, which is the strategy he will play. The same goes for the IA mechanism, where the only common best response strategy is to play rank-bias.

Under the pre-exam condition, Students 1 and 3 will both list three colleges. Suppose they believe College B will not participate with probability $p > 0$. Then they believe they play the matching game with three colleges with probability p, and with Colleges A and C with probability $1 - p$. It is easy to see that truth-telling is always the best response strategy for Student 1 under both mechanisms. For Student 3, listing $B - A - C$ is the only common best response for both scenarios. Therefore, both Students 1 and 3 will play the strategies under the game where there are three colleges.

To predict what will happen in this game environment, first we only look at the equilibrium strategies when there is centralized matching only, without the possibility of early admissions. Since under the DA mechanism under both timing conditions, truth-telling is the dominant strategy, and under IA-post, rank-bias is the equilibrium strategy once the test scores are revealed,[8] for all the three

[7] This can be thought of as the cost associated with market unraveling, for example, the cost of screening students individually; the cost of students not studying for the rest of their high school since they know they can go to colleges for sure; social costs of rent-seeking and corruption; etc.

[8] Note that the last-ranked student is indifferent among all strategies.

Table 1
Matching payoffs for students and colleges.

Student type	1	2	3	College type	A	B	C
Matched payoff	30	20	10	Matched payoff	30	20	10

Table 2
Exam outcomes for each student, $p = 0.5$.

Student	1	2	3	Prob.
Underperform	12	7	2	0.25
Normal	20	15	10	0.5
Overperform	28	23	18	0.25

mechanisms, matching outcomes are the same for any given realization of test scores. Table 3 summarizes the probability of ranking realizations of test scores and corresponding equilibrium outcomes. Most of the outcomes have blocking pairs and are unstable with respect to aptitude.

The choice of test score uncertainty p as well as the payoff choice of different matches jointly determine College A and B's expected payoffs in the centralized matching. Given the distribution of matching outcomes, College A will have lower risk-neutral expected utility (26.25) compared to the utility from its match under stable matching (30). It will have sufficient incentive to pay the 3-points fee and send an early admission offer to Student 1, and Student 1 will accept because accepting first-order stochastically dominates not doing so. Conditional on the strategies of College A and Student 1, College B will now have its own incentive to send an early admission offer to Student 2, and Student 2 will accept as well. Only College C will be left now and it admits Student 3. The outcome is now stable, but at a cost of 6 points out of 60. This is summarized in Tables 4 and 5.

However, IA-pre is a different story. Since students do not know the eventual realization of their test outcomes, they use strategies that maximize their expected utility. Again we only look at the equilibrium strategies for the centralized matching first. Under this game environment, students will play the following Bayesian Nash equilibrium strategies: Student 1 submits: $A - B - C$; Student B: $B - A - C$; and Student C: $B - A - C$. Note that Student 1 and 2 exhibit aptitude-based rank-bias strategies.[9] This Bayesian Nash equilibrium is unique and robust to risk attitude and parameter choices, as simulation using different combinations of parameters shows that equilibrium strategies do not change for $0.4 \leq p \leq 0.6$ and for risk-seeking and very risk-averse players.[10]

Given such strategies, College B will now have a strong incentive to send an early admission offer to Student 2, because Student 2 will accept it and Student 1 will not. Conditional on College B's play, College A will not have a strong enough incentive to send an offer to Student 1. Consequently, Student 1 goes to College A with probability 93.75%, and Student 3 6.25%. This is summarized in Tables 6 and 7.

4.2. Experiment procedure

Each session consists of 15 participants in 3 groups with 5 students in each group. Three participants of them play the role of the students, and two participants play the role of either College 1 or College 2. Since in equilibrium, College 3 never sends any offer, to reduce strategic uncertainty and control the beliefs of other players, it is played by the computer using the equilibrium strategy ("doing nothing"), which is common knowledge explained to the subjects in the instruction.

The experiment repeats for 20 rounds to facilitate learning. Each participant is assigned the role of either a college or a student at the beginning. Each round they are randomly rematched with different group members. Which college or student a participant plays is randomly decided each round. That is, a participant may play as different colleges or students, but he or she also plays as either a college or a student throughout the experiment.

During the early admission stage, we use the strategy method to elicit students' entire strategy profiles. Specifically, before students receive offers from colleges, they will be asked the following questions: "If you receive offers from both Colleges 1 and 2, what would you do? 1. Accept College 1's offer; 2. Accept College 2's offer; 3. Do not accept any offer.", "If you receive offer only from College 1, what would you do? 1. Accept the offer; 2. Do not accept the offer.", and "If you receive offer only from College 2, what would you do? 1. Accept the offer; 2. Do not accept the offer.".

At the end of the experiment, we elicit subjects' risk attitude using the lottery game from Holt and Laury (2002). The features of the experiment is summarized in Table A.6 in the Appendix.

The experiment was conducted in summer 2014 at the Experimental Economics Laboratory at The Ohio State University and the Behavioral and Experimental Economics Laboratory at the University of Michigan School of Information. There are 180 subjects across 12 sessions, with 3 sessions for each treatment. No one subject participated in more than one session. The average payment is $19.1, including a $5 show-up fee. Each session lasts about 90 min.

[9] Aptitude-based rank-bias happens to coincide with truth-telling for Student 1.

[10] This BNE does not change for risk parameter $0.3 \leq \gamma \leq 1.1$ with functional form $U(w) = \frac{w^\gamma}{\gamma}$.

Table 3
Possible ranking realizations after the exam.

Realized rankings	DA and post-exam IA NE	Probability	# Blocking pairs
1–2–3	1–2–3	0.406	0
2–1–3	2–1–3	0.266	1
1–3–2	1–3–2	0.266	1
3–1–2	3–1–2	0.016	2
3–2–1	3–2–1	0.031	3
2–3–1	2–3–1	0.016	2

Table 4
Colleges' incentives to make their move before the centralized matching begins when matching outcome is only stable w.r.t. test scores.

College	Exp. payoff post-exam	Payoff stable matching w.r.t. aptitude	Incentive
A	26.25	30	3.75
B	20	20	0
C	13.75	10	0
Total	60	60	

Table 5
Colleges' incentives to make their move before the centralized matching begins when matching outcome is only stable w.r.t. test scores, conditional on the equilibrium strategy of College A.

College	Exp. payoff post-exam	Payoff stable matching w.r.t. aptitude	Incentive
B	16.875	20	3.125
C	13.125	10	0
Total	30	30	

Table 6
Colleges' incentives to make their move before the centralized matching begins under pre-exam IA.

College	Exp. payoff pre-exam	Payoff stable matching w.r.t. aptitude	Incentive
A	30	30	0
B	16.875	20	3.125
C	13.125	10	0
Total	60	60	

Table 7
Colleges' incentives to make their move before the centralized matching begins under pre-exam IA, conditional on the equilibrium strategy of College B.

College	Exp. util. post-exam	Payoff stable matching w.r.t. aptitude	Incentive
A	28.75	30	1.25
C	11.25	10	0
Total	40	40	

5. Experimental results

Since the environment is two-sided, we report the behavior of colleges and students separately. In this section, we first report the decisions by colleges. Then we report the decisions by students. Finally, we report the matching outcomes, including matching stability with respect to both test scores and aptitude. Throughout this section, the general null hypothesis is that there is no difference in behaviors or matching outcomes.

5.1. Decisions by colleges and efficiency of the market

First, we examine the proportion of colleges who participate in the early admission process, as it directly represents market unraveling. Recall that as long as a college sends an offer to a student, the college pays a fixed amount of fee, regardless of whether that student accepts the offer or not eventually. Since without early admissions, the sum of payoff is the same for all matching outcomes, the proportion of colleges who send early admission offers therefore serves as the proxy for the efficiency of the market.

Fig. 2 shows the overall proportion of colleges who participate in early admissions as well as the proportions for College A and B separately. Theorem 1 and equilibrium for the game environment lead to the first hypothesis.

Hypothesis 1. (Participation in early admissions)**.** *There are significantly fewer colleges who participate in early admissions under the IA mechanism with pre-exam ROL submission than under the IA mechanism with post-exam submission or the Deferred Acceptance mechanism with either of the two timing conditions.*

Result 1. (Participation in early admissions)**.** *The IA mechanism with pre-exam submission performs best in reducing participation in early admissions. Specifically, the four treatments have the following order in occurrences of market unraveling (early admissions): IA-pre < DA-post < DA-pre < IA-post.*

Support. **.** *Two-sided proportion test shows that IA-pre has a significantly lower participation rate than the second best, DA-post (46.7%* vs *58.1%, $p = 0.002$). Similarly, DA-post has a lower rate than DA-pre (58.1%* vs *66.7%, $p = 0.017$), and DA-pre has a lower rate than IA-post (66.7%* vs *76.9%, $p = 0.002$).*

By Result 1, we reject the null in favor of Hypothesis 1. This finding confirms the theoretical prediction in Proposition 3 that mechanisms producing matchings unstable w.r.t. aptitudes will lead to market unraveling. It is also consistent with the equilibrium prediction for the experimental environment that IA mechanism, combined with pre-exam ROL submission, significantly reduces market unraveling.

We further break down the differences by colleges, and find that the difference between IA-pre and other three mechanisms becomes even larger for the top college, College A (46.1% vs 70.6% (second best), $p < 0.001$), and diminishes for the mid-tier college, College B (47.2% vs 40.6%, $p = 0.203$). This is also consistent with the equilibrium prediction of the experimental environment, where it predicts that College B under both mechanisms and both timing conditions would have incentives to participate in early admissions, while only under IA-pre it predicts that College A would not have enough incentive to do so.

In addition, Table 8 shows which students colleges send early admission offers to. Not surprisingly, the majority of College A's early admission offers are sent to Student 1, and College Bs send more offers to Student 2 than Student 1. However, there is still a sizable proportion of decisions deviating from the equilibrium predictions, particularly those of College B's. One possible explanation is strategic uncertainty. Since College B's optimal strategy is conditional on what early admission decision College A makes (as well as the possibility that students also do not play equilibrium strategies, especially in the case of IA-pre). Therefore, College B's decisions depend on what his belief about College A will do. If he believes that a sizable proportion of College A will not send early offers (in the case of IA-post, DA-pre and DA-post), then his expected incentive will very likely dip below the cost of 3 points.

Moreover, since the expected incentives in Tables 4 through 7 is calculated for risk neutral players, the deviations can also be explained by subjects' risk aversion. For example, in the experiment we find that about 43.9% of College A's nevertheless send out offers to Student 1. Since even under IA-pre, College A still has 6.25% chance of admitting the worst student, therefore, risk averse subjects may find it attractive to send out offers early and ensure himself of a match with a better student (given the equilibrium strategy of Student 1). In the case of College Bs, they might also participate in early admissions even though their expected incentives are smaller than the cost given their beliefs about College A's decisions. Regression confirms this conjecture (see Table A.3 and Fig. A.3 in the appendix): the more risk averse a College A player is, the more likely he or she will send out an early admission offer[11] (probit regression, marginal effect 11.4%, $p = 0.023$, standard errors clustered at the session level).

5.2. Students' strategies

In this section, we look at the decisions by students who form the other side of the market.

First we report students' decisions during the early admissions stage. Tables 9, 10 and 11 summarize the proportion of decisions when students receive both offers, the offer from only College A, and the offer from only College B, respectively. Consistent with the equilibrium prediction, an overwhelming majority of the students choose the offer from College A over the offer from College B. However, note that there are significant numbers of Student 1 s in all treatments that are also willing to accept the only offer from College B. This might be due to Student 1's failure to foresee non-equilibrium strategies by colleges (i.e. only College B but not College A sending them an early offer), therefore they choose carelessly their decisions regarding early offers from only College B. Also, under IA-post, there is a sizable proportion of Student B (proportion test against 0%, $p < 0.01$) who choose College B over College A (Table 9), even though that is a strictly dominated strategy. This is equivalent to rank bias, except that this happens during the early admissions. Such bias in this case is also unjustified.

The results of choices by students who are not admitted early (including truth-telling and rank-biased ROL submissions) are reported in Appendix A.4.

5.3. Matching outcomes

After looking at the decisions by both sides of the market, we now analyze the matching outcomes. The prediction for the proportion of students and colleges who are matched early is summarized in the following hypothesis. Fig. 3 shows the proportion of colleges who admit students early, while Fig. 4 shows the proportion of students who are admitted by colleges early. This helps to

[11] 23.3% of subjects who switch multiple times are excluded.

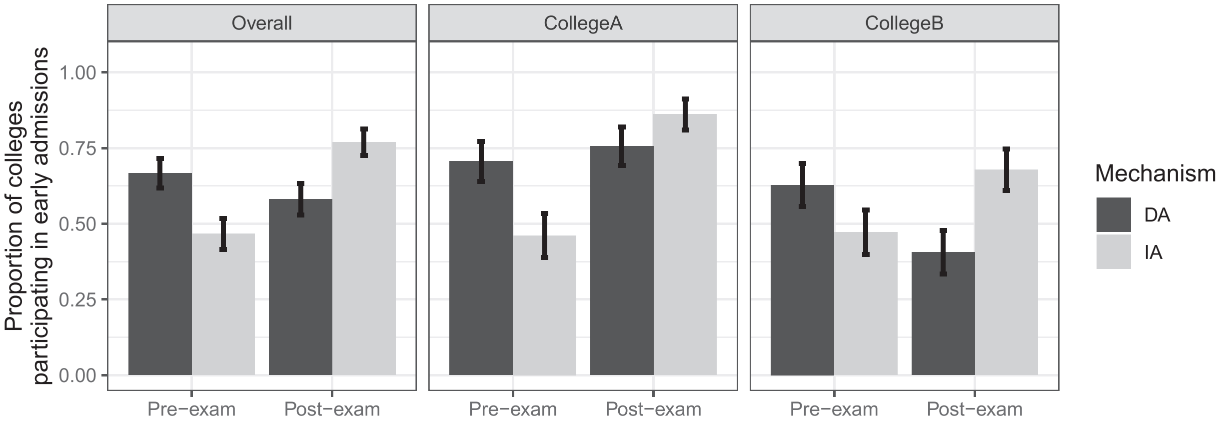

Fig. 2. Proportion of colleges who send out early admission offers (Error bars show 95% confidence intervals.)

Table 8
Percentage of students to whom offers are sent by colleges; equilibrium strategy in bold.

	IA		DA	
	College A	College B	College A	College B
	Pre-exam			
S1 (%)	43.9	16.1	**69.4**	21.1
S2 (%)	2.2	**24.4**	0.6	**39.4**
S3 (%)	0.0	6.7	0.6	2.2
None (%)	**53.9**	52.8	29.4	37.2
	Post-exam			
S1 (%)	**85.0**	12.2	**75.0**	11.7
S2 (%)	1.1	**55.0**	0.6	**27.8**
S3 (%)	0.0	0.6	0.0	1.1
None (%)	13.9	32.2	24.4	59.4

Table 9
Students' decisions when receiving both offers.

	IA			DA		
	S1	S2	S3	S1	S2	S3
	Pre-exam					
Accept A (%)	98.89	88.89	84.44	98.89	87.78	86.67
Accept B (%)	0.56	8.89	10.56	0.56	12.22	10.56
Reject (%)	0.56	2.22	5.00	0.56	0.00	2.78
	Post-exam					
Accept A (%)	92.2	77.8	75.6	99.4	91.1	81.7
Accept B (%)	3.9	20.0	20.0	0.0	8.9	16.1
Reject (%)	3.9	2.2	4.4	0.6	0.0	2.2

Table 10
Students' decisions when receiving offer from College A only.

	IA			DA		
	S1	S2	S3	S1	S2	S3
	Pre-exam					
Accept (%)	99.4	96.7	90.6	98.9	98.3	95.6
Reject (%)	0.6	3.3	9.4	1.1	1.7	4.4
	Post-exam					
Accept (%)	96.1	92.2	87.8	99.4	96.7	93.3
Reject (%)	3.9	7.8	12.2	0.6	3.3	6.7

Table 11
Students' decisions when receiving offer from College B only.

	IA			DA		
	S1	S2	S3	S1	S2	S3
	Pre-exam					
Accept (%)	41.1	93.3	95.6	47.8	78.9	96.7
Reject (%)	58.9	6.7	4.4	52.2	21.1	3.3
	Post-exam					
Accept (%)	46.7	83.3	92.2	32.2	75.6	97.8
Reject (%)	53.3	16.7	7.8	67.8	24.4	2.2

explain the empirical observation that after the policy change in Shanghai, more students were admitted early by elite colleges.

Hypothesis 2. (Early matchings)**.** *Fewer students and fewer colleges will be matched through early admissions under the IA mechanism with pre-exam ROL submission than under the IA mechanism with post-exam submission or the Deferred Acceptance mechanism with either of the two timing conditions.*

Result 2. (Early matchings)**.** *Fewer colleges admit students early under IA-pre, and fewer students are admitted early by colleges under IA-pre, than under each of the other three treatments.*

Support. **.** For colleges, two-sided proportion test shows that IA-pre has significantly lower early admission rate than the second best, DA-post (38.9% vs 49.2%, $p = 0.005$). DA-post is not significantly different from DA-pre (49.2% vs 53.3%, $p = 263$), and DA-pre has lower rate than IA-post (53.3% vs 65.3%, $p = 0.001$). For students, two-sided proportion test shows that IA-pre has significantly lower early admission rate than the second best, DA-post (25.9% vs 32.8%, $p = 0.013$). Similarly, DA-post is not significantly different from DA-pre (32.8% vs 35.6%, $p = 0.336$), and DA-pre has lower rate than IA-post (35.6% vs 43.5%, $p = 0.007$). Additional probit regressions with clustered standard errors confirm this finding, shown in Table A.6 and Table A.7 in the appendix.

By Result 2, we reject the null in favor of Hypothesis 2. The above result can be further broken down to different types of colleges and students. Generally, College As admit more students early than College Bs. Consistent with prediction, College As admit far fewer students early under IA-pre than the other three mechanisms. However, such difference does not exist for College Bs. Instead, College Bs admit most students under IA-post, and fewest under DA-post.

On the students' side, the proportion of students who are admitted early is directly related to the student quality of each type, and almost no Student 3 is admitted early. A pattern similar to colleges' early admission outcomes can be observed for students as well: far fewer Student 1 s are admitted early under IA-pre than the other three mechanisms, and more Student 2 s are admitted early under IA-post.

Next, we look at the matching outcomes in terms of the two criteria of stability: stability with respect to test score, and stability with respect to aptitude. Note that in the experimental game environment, through early admissions, the IA mechanism with post-exam submission and the Deferred Acceptance mechanism under both timing conditions completely eliminate instability with respect to aptitude. However, there is still a small chance (6.25%) of a matching outcome being unstable under the IA mechanism with pre-exam submission. The opposite is true for stability w.r.t. test score. Stability predictions are summarized in the following two hypotheses. Note that the predictions are the opposite of what would happen in the absence of early admissions. Fig. 5 shows the proportion of matching outcomes that satisfies each of the two stability notions.

Hypothesis 3. (Stability w.r.t. test score)**.** *When early admissions are allowed, the stability w.r.t. test score for the four treatments has the following order: IA-pre>IA-post=DA-pre=DA-post.*

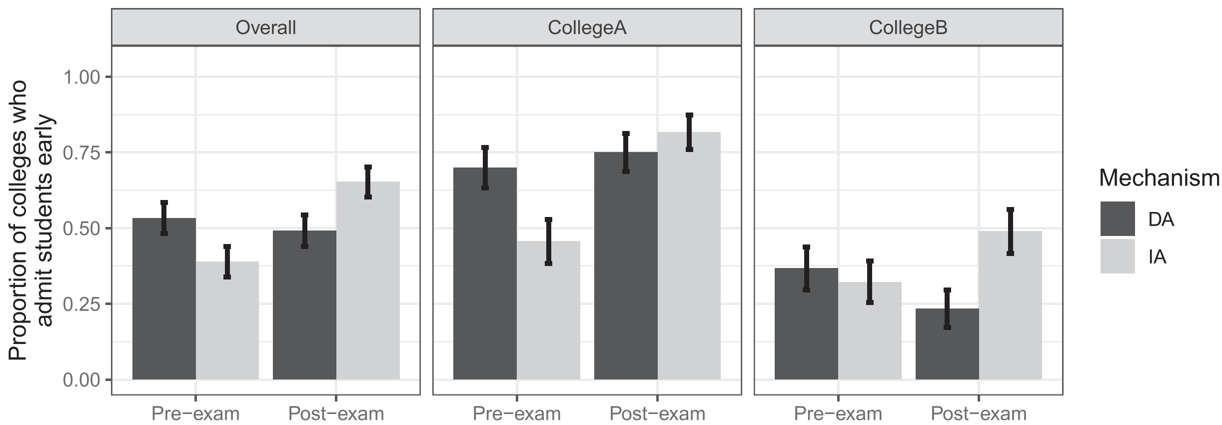

Fig. 3. Proportion of colleges who admit students early (Error bars show 95% confidence intervals.)

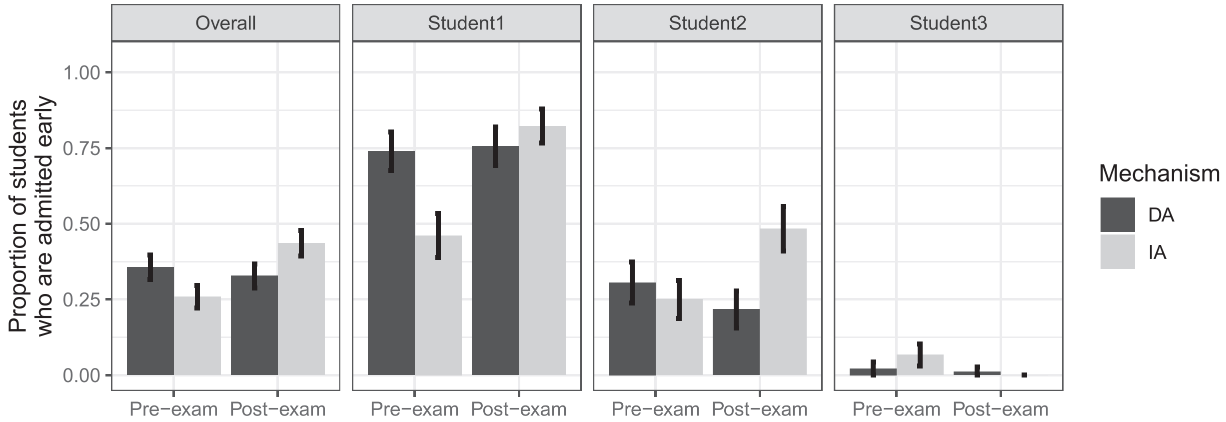

Fig. 4. Proportion of students who are admitted early (Error bars show 95% confidence intervals.)

Result 3. (Stability w.r.t. test score)**.** *DA is more stable than IA with respect to test score under both timing conditions. There is no difference within mechanisms between the two timing conditions.*

Support. . Two-sided proportion test shows that IA-pre has a significantly lower proportion of stable matchings w.r.t. test score than DA-pre (48.9% vs 61.1%, $p = 0.020$). Similarly, IA-post has a lower proportion than DA-post (50.0% vs 67.8%, $p < 0.001$). There is no timing effect ($p = 0.186$ and $p = 0.833$ for IA and DA respectively). Additional probit regressions with clustered standard errors confirm this finding, shown in Specification (1) in Table A.8 in the appendix.

Result 5 fails to reject the null hypothesis that there is no difference in stability (w.r.t. test score) within IA between pre and post timing conditions, when early admissions are present. It also contradicts the prediction that (1) IA-pre is more stable than DA-pre and (2) there is no difference across mechanisms under the post timing condition. These contradictions to theoretical prediction can be explained by colleges' failure to employ Nash equilibrium early admission strategies. Note that under DA, if colleges do not participate in early admissions, truth-telling by students leads to matching outcomes that are stable w.r.t. test scores. Table 8 in Section indeed shows that a significant proportion of college players deviating from the equilibrium strategies.

Next we explore stability with respect to aptitude.

Hypothesis 4. (Stability w.r.t. aptitude)**.** *When early admissions are allowed, the stability w.r.t. aptitude for the four treatments has the following order: IA-pre<IA-post=DA-pre=DA-post.*

Result 4. (Stability w.r.t. aptitude)**.** *IA-post is more stable with respect to aptitude than IA-pre. There is no difference between IA-pre and DA-pre. There is also no difference between the two treatments under DA.*

Support. . Two-sided proportion test shows that IA-pre has a significantly lower proportion of stable matchings w.r.t. aptitude than IA-post (62.8% vs 73.9%, $p = 0.024$). However, there is no significant difference between IA-pre and DA-pre (62.8% vs 66.7%, $p = 0.440$), IA-post and DA-post (73.9% vs 67.8%, $p = 0.202$), and DA-pre and DA-post (66.7% vs 67.8%, $p = 0.822$). Additional probit regressions with clustered standard errors confirm this finding, shown in Specification (2) in Table A.8 in the appendix.

From the above result, although we can reject the null that there is no timing effect under IA, there is no significant mechanism

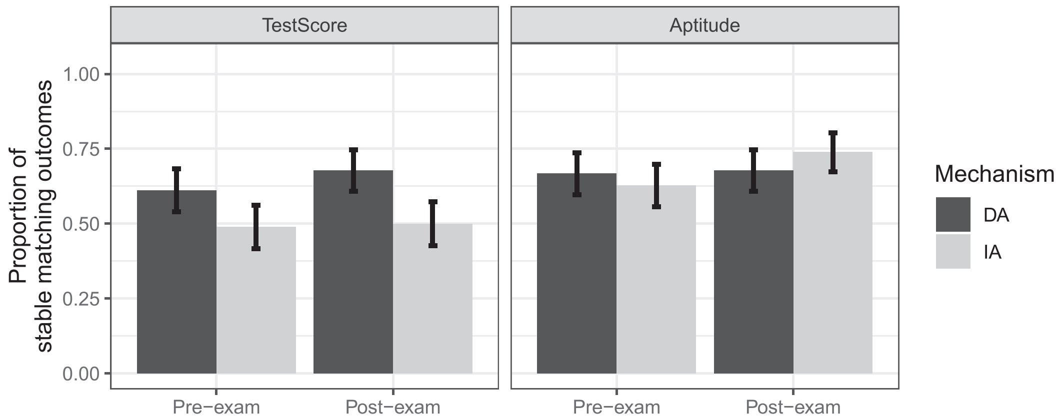

Fig. 5. Proportion of stable matchings w.r.t. test score (left) and w.r.t. aptitude (right) (Error bars show 95% confidence intervals.)

difference under the pre condition. For the rest of this section, we explore why this is the case. Note that in order to reach predicted matching outcomes, both colleges and students need to play Nash equilibrium strategies. However, from previous finding we see that this is not the case. Therefore, we look at how the stable matchings (w.r.t. aptitude) are achieved, and why some matching outcomes are unstable (w.r.t. aptitude).

First we investigate how stable matchings are achieved. Fig. 6 presents the proportion of stable matchings (w.r.t. aptitude) that are achieved without any early admissions under the four mechanism-timing conditions we study.

Result 5. . *More stable matchings w.r.t. aptitude are achieved without going through any early admissions under IA-pre.*

Support. . Two-sided proportion test shows that IA-pre has a significantly higher proportion of stable matchings (w.r.t. aptitude) achieved without early admissions, compared to the second best, DA-pre (28.3% vs 14.2%, $p = 0.008$). There is no significant difference in proportion between DA-pre and DA-post (14.2% vs 11.5%, $p = 0.531$). IA-post has the lowest proportion (3.0% vs 11.5%, $p = 0.008$). Additional probit regressions with clustered standard errors confirm this finding, shown in Specification (3) in Table A.8 in the appendix.

The above result is encouraging: we were worried that, per equilibrium prediction, IA-pre trades more early admissions for more unstable matching outcomes w.r.t. aptitude. Now we see that the difference in stability is so small, especially when compared to the reduction in market unraveling, that the benefits outweigh the costs.

One may argue that by deliberately allowing early admissions as a policy choice, even though it is costly, it can solve the problem of aptitude mismatch ex post. The above result, however, shows that simply through the design of the centralized matching, similar outcomes can be achieved, for a much lower cost. Of course, if the top concern is stability w.r.t. test score, regardless of timing conditions, DA is still the preferred choice; then it is probably even better to simply restrict the use of early admissions.

Finally, we look at matching outcomes that are unstable w.r.t. aptitude rising under different matching processes. Table 12 presents the classification of causes of unstable matchings. We identify three categories of causes: (1) no student participates in early admissions (thus all three participate in the centralized matching), and measurement error of the test causes unstable matchings; (2) one student participates in early admissions (thus two participate in the centralized matching), the early matching results are stable, and measurement error of the test causes unstable matchings; and (3) colleges send "wrong" offers to students, therefore early admission itself directly leads to unstable matchings. This table shows that IA-pre has the highest occurrence of unstable matchings caused by the measurement error in the test with all students participating in the centralized matching. However, it has the lowest occurrence of unstable matchings when one student is admitted through early admissions. Additionally, IA-pre again has the highest number of cases of unstable matchings caused by colleges admitting early worse students than the students they could have got under the equilibrium. Many of such cases can be avoided by not participating in the early admission at all.

5.4. Centralized matching without early admissions

Analyzing matching stability with early admissions has one main shortcoming: we do not know how stable matching outcomes would have been without the possibility of early admissions; therefore, we do not have an empirical measure of whether colleges' responses to potential market instability are justified. Moreover, we also do not know the effect of allowing early admissions on market stability. Therefore, we conducted another round of experiments using the same design parameters, except there is no early admission stage and colleges have no say in the admission outcomes. There are a total of 96 subjects with 24 in each of the treatment. The experiment was conducted at Shanghai Jiao Tong University Smith Experimental Economics Research Center in November 2022. Subjects only play the role of students; the game is repeated for 20 rounds with random rematching each round. We would also advise caution when drawing direct comparison with the original experiment since the new experiment was conducted at a different time with a different subject pool.

The first observation we have (illustrated in Fig. 7) is that with only centralized matching, IA-pre, as expected, has the lowest stability w.r.t. scores. It also has the highest proportion of stable matching w.r.t. aptitude at 48.75%. While the difference is not significant from pre-exam DA, it is significantly higher than both mechanisms under post-exam timing condition ($p = 0.036$ vs. IA-post, $p = 0.07$ vs. DA-post, one-sided proportion test). Compared with the original experiment, we also see a strong "correction effect" of early admission: none of the new treatments has a stability w.r.t. aptitude exceeding 50%, while all original treatments have this measure above 60%. We also observe an obvious trade-off between the two types of stability, consistent with theoretical prediction.

We also calculate the distance between the actual realized payoffs in the experiment and the theoretical stable matching payoffs when early admissions are available, which measures the (risk neutral) incentives for Colleges A and B, shown in Fig. 8. Consistent with the theoretical predictions in Section 4, the payoff distances for College A's in both mechanisms post-exam all surpass the predicted 3.75. The distance under DA-pre, while only close to the cost threshold of 3 utils, is still significantly higher than that under IA-pre ($p = 0.012$ one-sided t-test). The incentive for College B under IA-pre also surpasses the predicted 3.125. This result shows that the decisions by colleges to participate in the early admissions are justified, as the cost of not doing so outweighs the cost of early admission itself.

6. Conclusion

Using standardized tests to evaluate students and select them into higher levels of education has become a hot topic around the world, whether it being a long used practice or not. One of the main problems with standardized tests, however, is the associated measurement errors. Therefore, the misalignment of priorities and preferences in the matching process is almost inevitable. Although

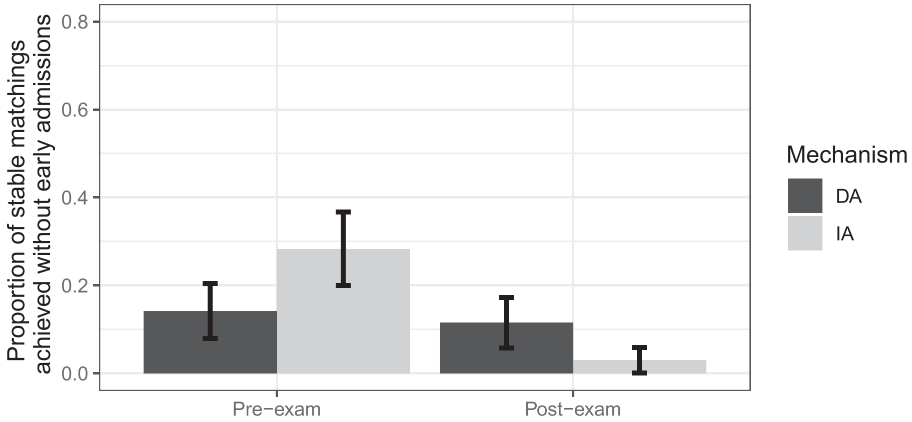

Fig. 6. Proportion of stable matchings achieved without early admissions (Error bars show 95% confidence intervals.)

Table 12
Matching outcomes that are unstable w.r.t. aptitude under different matching processes.

	Pre-exam		Post-exam	
	IA	DA	IA	DA
No early matching; Unstable centralized matching	46.3%	25%	29.8%	29.3%
Stable early matching; Unstable centralized matching	22.4%	51.7%	57.4%	62.1%
Unstable early matchings	31.3%	23.3%	12.8%	8.5%
Total cases	67	60	47	58

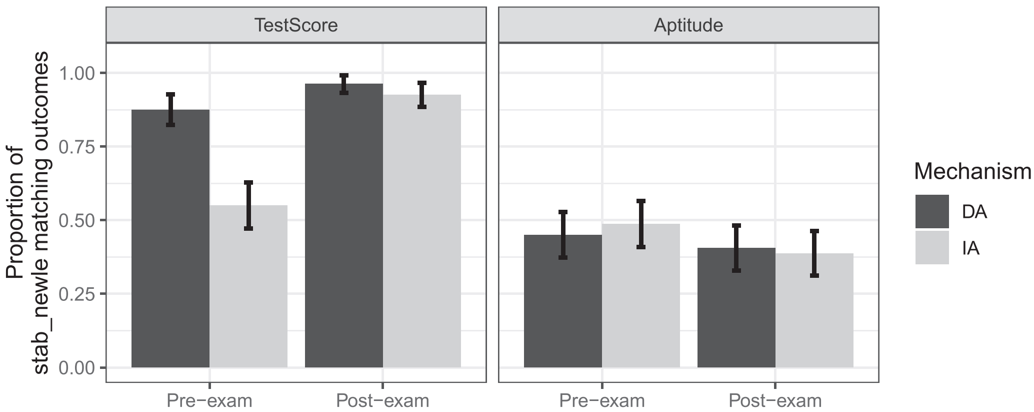

Fig. 7. Proportion of stable matchings w.r.t. test score (left) and w.r.t. aptitude (right) without early admissions (Error bars show 95% confidence intervals.)

stable matching mechanisms such as the Deferred Acceptance mechanism are generally considered superior to unstable alternatives, especially in reducing market unraveling, this may not be the case under such case. In this paper, motivated by an unraveling market in the real world, we try to explain why a transition to an apparently more stable matching mechanism increases rather than decreases the level of instability in this market.

We find that under the context of college admissions, when students' priorities in the matching are determined by test scores from a standardized test, which is a noisy realization of students' aptitudes, any matching mechanism that produces outcomes that are stable with respect to test score, produces outcomes that are not stable with respect to aptitude. If colleges only participate in the centralized admission process, then they may prefer students who are matched with other colleges, and these students also prefer to be matched with these colleges rather than the colleges they are currently matched with. This gives colleges incentives to circumvent the centralized matching, and try to admit these students in advance. Since it is the dominant strategy to reveal one's true preference over colleges under the Deferred Acceptance mechanism, the matching respects test scores absolutely, with no regard to a student's actual aptitude. Therefore, this mechanism will cause market unraveling through early admissions by colleges. On the contrary, the IA mechanism is manipulable, and the information of students' rankings, either of test scores or of aptitudes, is used by students to

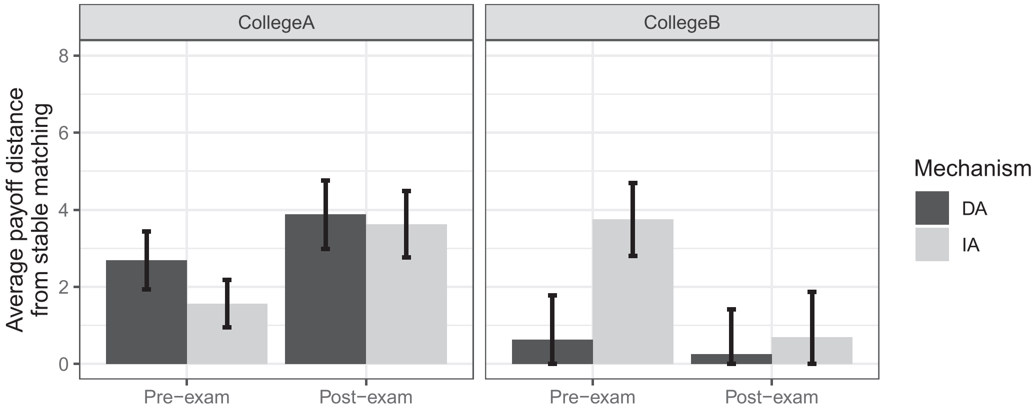

Fig. 8. Average payoff distance from stable matchings for College A (left) and College B (right) without early admissions (Error bars show 95% confidence intervals.)

strategize their rank-ordered lists. Therefore, we can take advantage of such manipulability by revealing the set of information that is desirable (in this case, the rankings of aptitudes rather than the rankings of test scores).

To test the theoretical predictions, we design a laboratory matching market, where colleges can send out early admission offers to students ahead of the centralized matching. Yet, they can not actively participate in the matching itself once it starts. Experimental results show that, consistent with theoretical predictions as well as real world evidence, the Deferred Acceptance mechanism causes more market unraveling as more colleges send out early admission offers, circumventing the centralized matching. However, when students are required to submit their rank-ordered lists before they take the exam, when they only know their rankings in aptitudes, the IA mechanism significantly reduces the occurrence of early offers. Furthermore, the IA mechanism with pre-exam submission does not yield more unstable matching outcomes with respect to students' aptitudes compared to other mechanisms, even though fewer students and colleges go through early admissions. Even for the stable matching outcomes, significantly more of them are achieved without anyone participating in early admissions. Thus, we conclude that the IA mechanism under pre-exam submission condition performs at least as well as all other mechanisms and timing conditions in stability with respect to aptitude, and it achieves that with much less sacrifice in market efficiency caused by unraveling.

While this paper shows that the IA mechanism can outperform the Deferred Acceptance mechanism in certain measures, we do not attempt to make policy recommendation of some mechanisms over some others, since although reducing market unraveling can be important, it can be equally important to some that instability with respect to test score be avoided as much as possible, due to political, legal, or other reasons. There are, however, ways to reconcile the conflict between the two types of (in)stability, as our results imply. One way is to reduce the measurement errors of college entrance exams, as it directly leads to instability w.r.t. aptitude. Although more costly, it can be achieved through multiple testings. For example, in Shanghai, high school students can take English language part of the college entrance exam twice a year, significantly reducing score uncertainty. The second way is to allow but regulate early admissions, aiming to reduce the financial and social cost of the practice. This can be done through the standardization of early admission timeline and procedures.

Data availability

Data will be made available on request.

Acknowledgements

The authors thank Yan Chen, Tayfun Sönmez, YingHua He, and Stephen Leider for valuable comments and suggestions. The authors also thank Paul J. Healy and Anthony Bradfield for help with the access to The Ohio State University Experimental Economics Laboratory. Jiang thanks the National Natural Science Foundation of China (Grant 72103135 and Grant 72231003) for research support. Li thanks the National Social Science Fund of China (Grant 21AZD036) and the National Natural Science Foundation of China (Grant 72274062) for research support.

Appendix A. Appendix

A.1. Theoretical results of rank bias under IA

Proposition 4. Rank bias is a Nash equilibrium strategy under both the IA and the DA mechanisms with post-exam ROL submission.

Proof. . Given the definition of rank bias, the mass of students who list college c_j as their first choices using rank bias strategy equals to q_{c_j}. This also implies that all students get into their first choice colleges.

No student has the incentive to deviate from that strategy. Students who are ranked top q_{c_1} will not deviate because they will certainly be admitted by the top college. Given that, students who are ranked between $q_{c_1}+1$ and q_{c_2} cannot profitably deviate as well: if they rank a college that is below the second best college, they are strictly worse off; if they rank a college that is above the second best college, their chance of getting into that is 0. The same reasoning can be applied to all students down the rank. Finally for the bottom ranked students, they are indifferent between playing rank bias or ranking any college as their first choice.

A.2. Proof of Theorem 1

Proof. . The proof consists of the following steps.

First, we show that the admissions results are assortative. That is, top q_{c_1} students in the test are admitted into college c_1, next q_{c_2} students are admitted into college c_2, etc. This is evident given the truth-telling property of the DA mechanism and the homogeneous preferences of students.

Denoting the density function of test scores as $h(\cdot)$, we partition students' test scores into m parts with cut-off points $p_{t_1}, \ldots, p_{t_{m-1}}$, such that

$$\int_{p_{t_{i-1}}}^{p_{t_i}} h(x)dx = \frac{q_{c_i}}{\sum q_c}.$$

From the above result we know that a student with test score t, $p_{t_{i-1}} \le t < p_{t_i}$ is admitted to college c_i. Note that for colleges that are the boundary (best school and worst school), we replace cut-off points with $-\infty$ and ∞.

We do the same to aptitude so that we have cut-off points $p_{a_1}, \ldots, p_{a_{m-1}}$. A student with score-aptitude pair (t,a) admitted by a non-boundary college c_i will form a block pair with at least one better college if his aptitude satisfies $a > q_{a_i}$. The probability this will happen is given by: $Prob(a > q_{a_i} | q_{t_{i-1}} \le t < q_{t_i})$, which is expanded to

$$Prob\left(a > p_{a_i} | p_{t_{i-1}} \le t < p_{t_i}\right) = \frac{Prob\left(a > p_{a_i}, p_{t_{i-1}} \le t < p_{t_i}\right)}{Prob\left(p_{t_{i-1}} \le t < p_{t_i}\right)}$$

We know that $t = a + \eta$, where a has a p.d.f. f and η has a p.d.f. g. We need to calculate the joint p.d.f. of $(a, a+\eta)$. Let $X = a$, $Y = a + \eta$. Solving the linear equation, we have $a = X$, $\eta = Y - X$. We calculate the determinant of the Jacobian matrix:

$$\begin{vmatrix} \frac{da}{dX} & \frac{da}{dy} \\ \frac{d\eta}{dX} & \frac{d\eta}{dy} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1$$

So the joint distribution of $(a, a+\eta)$ is given by $f_{X,\,Y}(x,y) = f(x)g(y-x)$. Go back to the probability equation, we have:

$$Prob\left(a > p_{a_i} | p_{t_{i-1}} \le t < p_{t_i}\right) = \frac{\int_{p_{a_i}}^{\infty} \int_{p_{t_{i-1}}}^{p_{t_i}} f(x)g(y-x)dydx}{\int_{p_{t_{i-1}}}^{p_{t_i}} \int_{-\infty}^{\infty} f(x)g(z-x)dxdz} > 0$$

given the non-trivial support of g.

A.3. Proof of Proposition 3

Proof. . Given the assortative nature of the game environment, we first consider the decision of the top ranked college c_1. It has the incentive to send offers and admit early the top q_{a_1} of students. The incentive is determined by the proportion of all students that will form a blocking pair with c_1:

$$| S | \cdot Prob\left(a > p_{a_1} | t < p_{t_1}\right) = | S | \cdot \frac{\int_{p_{a_1}}^{\infty} \int_{-\infty}^{p_{t_1}} f(x)g(y-x)dydx}{\int_{-\infty}^{p_{t_1}} \int_{-\infty}^{\infty} f(x)g(z-x)dxdz}$$

using the same notation as in Theorem 1. Such incentive is independent of the decisions of lower ranked colleges, because any offers sent will be accepted given the homogeneity of students' preferences over colleges.

Assuming the rest of the colleges and students correctly anticipate c_1's decision and outcome, the resulting market then essentially becomes the original market with c_1 and top q_{a1} students removed, where c_2 and q_{a_2} are now top ranked instead. Apply the above step repeatedly until we have one college (the original bottom ranked) left in the market, who no longer will form any blocking pairs with students.

A.4. Results of student ROL submissions during centralized matching

The following hypothesis comes from the theoretical prediction of the truth-telling property of the DA and IA mechanism. Fig. A.1

reports the proportion of truthful preference revelation among students who are not admitted early and therefore participate in the main admissions (note that students who are admitted early do not make these decisions). We can see clear difference in the proportions of truth-telling between IA and DA.

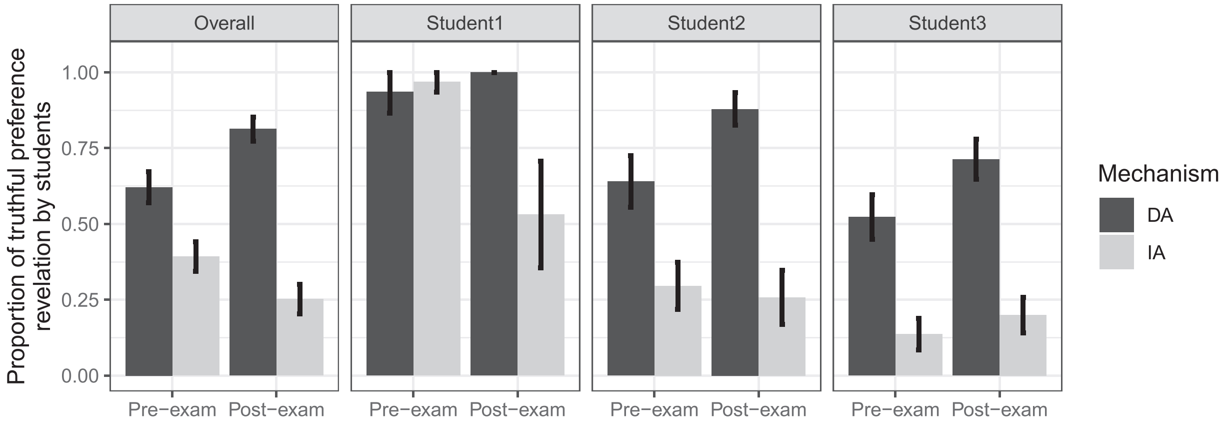

Fig. A.1. Proportion of truthful preference revelation by students (Error bars show 95% confidence intervals.)

Hypothesis A.1. (Truthful preference revelation). *More students will reveal their true preferences over colleges under the Deferred Acceptance mechanism than under the IA mechanism.*

Result A.1. (Truthful preference revelation). *Conditional on not being admitted early, significantly more students report their true preferences of colleges under DA than under IA.*

Support. . Two-sided proportion test shows that IA-pre has significantly lower truth-telling rate than DA-pre (39.3% vs 62.1%, $p < 0.001$). Similarly, IA-post has lower rate than DA-post (25.5% vs 81.3%, $p < 0.001$). Additional probit regressions with clustered standard errors confirm this finding, shown in Table A.5 in the appendix.

The above result rejects the null hypothesis that there is no significant difference in truth-telling between the two mechanisms, in favor of Hypothesis A.1. In addition to the mechanism effect on truth-telling, we also observe the effect of timing. Significantly more students reveal their true preferences under DA-post than under DA-pre (proportion test, 81.3% vs 62.1%, $p < 0.001$). Further investigation shows that this is mostly attributed to the misrepresentations by Student 2 and Student 3.

On the other hand, the difference in truth-telling under IA is mostly attributed to the top ranked student, Student 1. Many fewer Student 1 s reveal their true preferences under IA-pre compared to each of the other three treatments. This is not surprising because under IA, when a student knows that he no longer ranks at the top after the test, the best strategy is not to list the best college as his first choice, but rather the college that corresponds to his ranking instead.

Hypothesis A.2 is derived from the predicted student strategies in Section 4.

Hypothesis A.2. (ROL strategies). *More students will rely on their rankings of test scores when submitting ROLs under the IA mechanism with post-exam submission, while more students will rely on their rankings of aptitude under the IA mechanism with pre-exam submission.*

Result A.2. (ROL strategies). *Students exhibit rank bias under IA with post-exam ROL submission, while they do not with pre-exam ROL submission.*

Support. . Specifications (1) to (4) in Table A.1 represent the effect of ranks in different treatments. Under the pre-exam timing conditions ((1) and (3)), rank in test score has no significant effect. Under IA-post, having a higher rank in the test leads to a significantly higher chance of listing a better college as first choice. Surprisingly, we also observe rank-bias under DA-post, although the effect is smaller than under IA-post.

Table A.1
Order probit: Rank of test score on Student 2's first choice.

Dep. Var.	First choice college			
	(1)	(2)	(3)	(4)
	IA-pre	IA-post	DA-pre	DA-post
Rank at FirstChoice=1	0.021	−0.252***	0.092	−0.148***
	(0.051)	(0.043)	(0.061)	(0.036)
Rank at FirstChoice=2	−0.019	0.105**	−0.071*	0.098***
	(0.048)	(0.052)	(0.040)	(0.031)

(*continued on next page*)

Table A.1 (*continued*)

Dep. Var.	First choice college			
	(1)	(2)	(3)	(4)
	IA-pre	IA-post	DA-pre	DA-post
Rank at FirstChoice=3	−0.001	0.147***	−0.021	0.051*
	(0.003)	(0.054)	(0.027)	(0.028)
Observations	135	93	125	141

Notes: Standard errors in parentheses are clustered at the session level. Reporting marginal effects for different outcomes. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.1 reports four ordered-profit specifications testing the effect of rank in test score on Student 2's first choice. Student 2 is the most interesting target of this task, because most Student 1 s are admitted early, and Student 3 is indifferent among multiple strategies.

By Result A.2, we reject the null in favor of Hypothesis A.2. In addition to what is stated in the hypothesis, we also observe that submission timing even has an effect under DA as well, which is consistent with the district bias findings in experimental school choice literature (e.g. Chen and Sönmez (2006); Chen and Kesten (2019)), where students exhibit district school (high-priority) bias under DA.

A.5. Additional figures

The following graph goes into what actually happened behind each dot in the previous graph by showing the distribution of aptitudes of students admitted by each college in a single simulation. The simulated matching has a standardized test with standard deviation 8. The distribution curves from left to right are for Colleges 1, 2, 3, 4, and 5, in order. Vertical lines from left to right are 20%, 40%, 60%, and 80% quantiles for students' aptitudes. In a matching that is stable with respect to aptitude under the simulated environment, the top 20% of students should be matched with College 1, the next 20% with College 2, etc. Therefore, for College 1, the area to the right of the 80% quantile line that is a part of the distribution of aptitudes of students admitted by colleges of lower quality is the proportion of students College 1 forms blocking pairs with. The same applies to Colleges 2 through 5. Fig. A.2 illustrates the pattern of unstable mismatch with respect to aptitude through overlapped areas under the aptitude distributions, which are in turn caused by the measurement error of the standardized test.

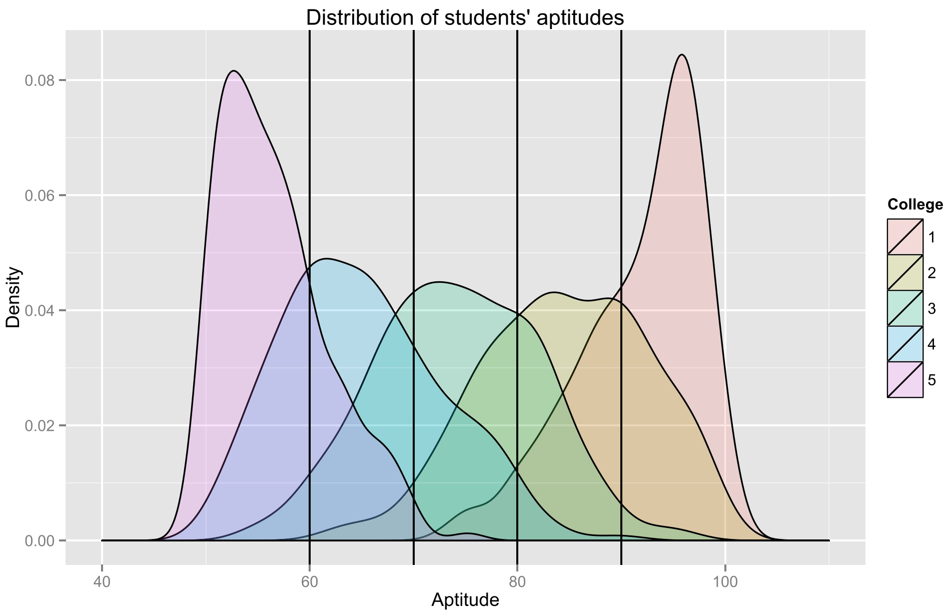

Fig. A.2. Number of blocking pairs increases with measurement error.

The following figure shows the relationship between risk attitude, measured by the switching points in the lottery game, and the likelihood of participating in early admissions for college players (the larger the switch point is, the more risk averse a subject is).

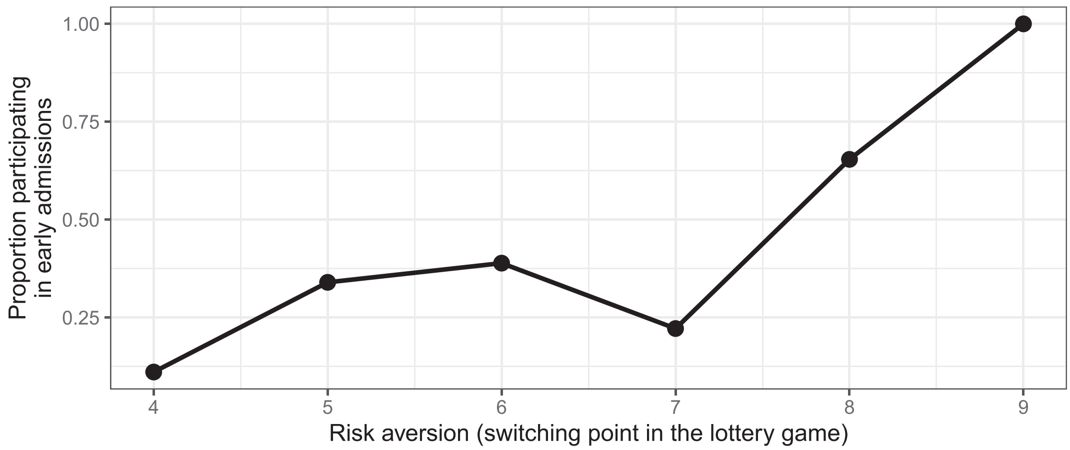

Fig. A.3. Relationship between risk attitude and likelihood of participating in early admissions.

The following graph shows College A's and Bs decision to participate in early admissions over time under different treatment.

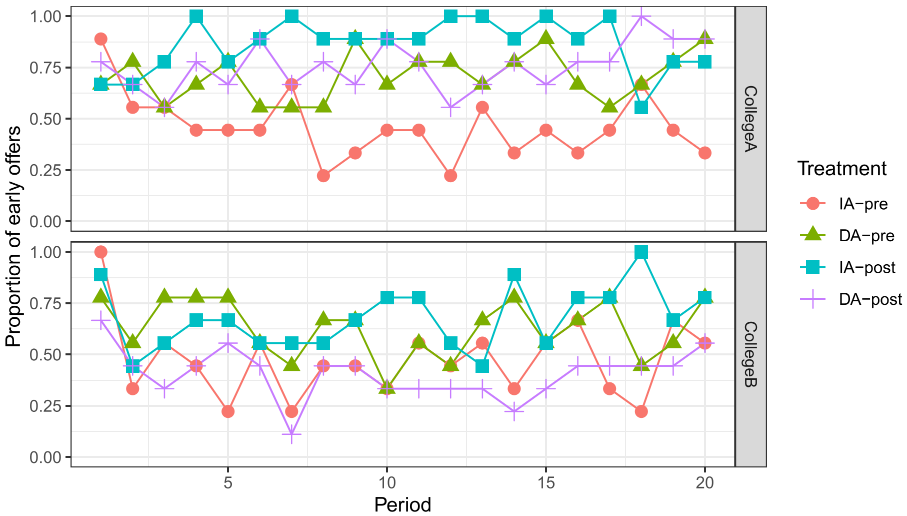

Fig. A.4. Proportion of colleges participating in early admissions over 20 experimental periods.

A.6. Additional tables for experimental design

Table A.2
Design features of the experiment.

	Pre-exam
IA	Student and colleges participate in early admissions Students submit ROLs Test scores are revealed Matches are calculated using the IA mechanism Risk preference elicitation at the end
DA	Student and colleges participate in early admissions Students submit ROLs Test scores are revealed Matches are calculated using the DA mechanism Risk preference elicitation at the end
	Post-exam
IA	Student and colleges participate in early admissions Test scores are revealed Students submit ROLs Matches are calculated using the IA mechanism

(*continued on next page*)

Table A.2 (*continued*)

	Pre-exam
DA	Risk preference elicitation at the end Student and colleges participate in early admissions Test scores are revealed Students submit ROLs Matches are calculated using the DA mechanism Risk preference elicitation at the end

Notes: During the early admission stage, colleges choose whether to pay a fee to participate in early admissions; if they choose to, they then decide which student to send offers to; simultaneously, students decide, if they receive any offers, whether or which to accept.

A.7. Additional regression tables

Table A.3
Probit regression: Risk preference on early admission decision.

Dep. Var.	Early admission
	(1)
HL switch point	0.114**
	(0.050)
Observations	138

Notes: Standard errors in parentheses are clustered at the session level; reporting marginal effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.4
Probit regression: Treatment effects on early admission decision.

Dep. Var.	Early admission		
	(1)	(2)	(3)
Post	0.299***	0.421***	0.421***
	(0.079)	(0.076)	(0.076)
DA	0.187*	0.229**	0.229**
	(0.108)	(0.115)	(0.115)
Post×DA	−0.381***	−0.373***	−0.373***
	(0.135)	(0.132)	(0.132)
CollegeB		0.015	0.015
		(0.028)	(0.028)
Post×CollegeB		−0.240***	−0.240***
		(0.057)	(0.057)
DA×CollegeB		−0.095*	−0.095*
		(0.053)	(0.053)
Period			0.001
			(0.002)
Observations	1440	1440	1440

Notes: Standard errors in parentheses are clustered at the session level; reporting marginal effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.5
Probit regression: Treatment effects on truth-telling.

Dep. Var.	Truth-telling		
	(1)	(2)	(3)
Post	−0.134***	−0.345***	−0.338***
	(0.020)	(0.103)	(0.104)
DA	0.197***	0.151	0.159
	(0.026)	(0.132)	(0.128)
Post×DA	0.331***	0.190***	0.186***
	(0.059)	(0.065)	(0.064)
Student2		−0.609***	−0.606***
		(0.051)	(0.052)

(*continued on next page*)

Table A.5 (*continued*)

Dep. Var.	Truth-telling		
	(1)	(2)	(3)
Post×Student2		0.349***	0.346***
		(0.102)	(0.101)
DA×Student2		0.130	0.123
		(0.111)	(0.108)
Student3		−0.715***	−0.714***
		(0.061)	(0.059)
Post×Student3		0.348***	0.343***
		(0.117)	(0.116)
DA×Student3		0.112	0.106
		(0.129)	(0.126)
Period			0.004*
			(0.002)
Observations	1416	1416	1416

Notes: Standard errors in parentheses are clustered at the session level; reporting marginal effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.6
Probit regression: Treatment effects on early matchings for colleges.

Dep. Var.	Matched early		
	(1)	(2)	(3)
Post	0.261***	0.347***	0.346***
	(0.068)	(0.064)	(0.064)
DA	0.142*	0.216*	0.216*
	(0.107)	(0.114)	(0.114)
Post×DA	−0.302**	−0.291**	−0.290**
	(0.134)	(0.133)	(0.133)
College2		−0.125***	−0.125***
		(0.012)	(0.012)
Post×College2		−0.193***	−0.193***
		(0.019)	(0.019)
DA×College2		−0.171***	−0.171***
		(0.018)	(0.018)
Period			0.006**
			(0.002)
Observations	1440	1440	1440

Notes: Standard errors in parentheses are clustered at the session level; reporting marginal effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.7
Probit regression: Treatment effects on early matchings for students.

Dep. Var.	Matched early		
	(1)	(2)	(3)
Post	0.175***	0.227***	0.227***
	(0.046)	(0.041)	(0.041)
DA	0.100*	0.161**	0.160**
	(0.071)	(0.078)	(0.078)
Post×DA	−0.202**	−0.196**	−0.195**
	(0.093)	(0.092)	(0.092)
Student2		−0.141***	−0.141***
		(0.014)	(0.014)
Post×Student2		−0.086***	−0.086***
		(0.028)	(0.028)
DA×Student2		−0.132***	−0.133***
		(0.024)	(0.025)
Student3		−0.381***	−0.381***
		(0.078)	(0.078)
Post×Student3		−0.364**	−0.365**
		(0.148)	(0.148)
DA×Student3		−0.193	−0.193
		(0.118)	(0.118)
Period			0.004**
			(0.002)
Observations	2160	2160	2160

Notes: Standard errors in parentheses are clustered at the session level; reporting marginal effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.8
Probit regression: Treatment effects on stability.

Dep. Var.	Stable w.r.t.	Stable w.r.t.	Stable
	Test score	Aptitude	w/o early admission
	(1)	(2)	(3)
Post	0.011	0.112*	−0.262***
	(0.056)	(0.058)	(0.079)
DA	0.119**	0.037	−0.100
	(0.053)	(0.072)	(0.087)
Post×DA	0.058	−0.101	0.236**
	(0.090)	(0.085)	(0.100)
Observations	720	720	488

Notes: Standard errors in parentheses are clustered at the session level; reporting marginal effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The dependent variable for specification (1) is whether a matching outcome is stable w.r.t. test scores; for specification (2) is whether a matching outcome is stable w.r.t. aptitude; and for specification (3) is whether a stable matching outcome is not achieved through early admissions.

Appendix B. Experimental instructions

Please see the supplementary material for the complete experimental instructions.

Appendix C. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chieco.2023.101963.

References

Abdulkadiroğlu, A., Che, Y.-K., & Yasuda, Y. (2011). Resolving conflicting preferences in school choice: The Boston mechanism reconsidered. *American Economic Review, 101*(1), 399–410.
Abdulkadiroğlu, A., Che, Y.-K., & Yasuda, Y. (2015). Expanding "choice" in school choice. *American Economic Journal: Microeconomics, 7*(1), 1–42.
Abdulkadiroğlu, A., Pathak, P. A., & Roth, A. E. (2005). The New York City high school match. *American Economic Review, 95*(2), 364–367.
Abdulkadiroğlu, A., Pathak, P. A., Roth, A. E., & Sönmez, T. (2005). The Boston public school match. *American Economic Review, 95*(2), 368–371.
Abdulkadiroğlu, A., & Sönmez, T. (1999). House allocation with existing tenants. *Journal of Economic Theory, 88*, 233–260.
Abdulkadiroğlu, & Sönmez, T. (2003). School choice: A mechanism design approach. *American Economic Review, 93*(3), 729–747.
Avery, C., Fairbanks, A., & Zeckhauser, R. J. (2009). *The Early Admissions Game.* Harvard University Press.
Avery, C., & Levin, J. (2010). Early admissions at selective colleges. *American Economic Review, 100*(5), 2125–2156.
Balinski, M. L., & Sönmez, T. (1999). A tale of two mechanisms: Student placement. *Journal of Economic Theory, 84*, 73–94.
Bloom, B. S., Madaus, G. F., Hastings, J. T., et al. (1981). *Evaluation to Improve Learning.* McGraw-Hill.
Chen, Y., Jiang, M., & Kesten, O. (2020). An empirical evaluation of chinese college admissions reforms through a natural experiment. *Proceedings of the National Academy of Sciences, 117*(50), 31696–31705.
Chen, Y., & Kesten, O. (2017). Chinese college admissions and school choice reforms: A theoretical analysis. *Journal of Political Economy, 125*(1), 99–139.
Chen, Y., & Kesten, O. (2019). Chinese college admissions and school choice reforms: An experimental study. *Games and Economic Behavior, 115*, 83–100.
Chen, Y., & Sönmez, T. (2002). Improving efficiency of on-campus housing: An experimental study. *American Economic Review, 92*(5), 1669–1686.
Chen, Y., & Sönmez, T. (2006). School choice: An experimental study. *Journal of Economic Theory, 127*(1), 202–231.
Dargnies, M.-P., Hakimov, R., & Kübler, D. (2019). Self-confidence and unraveling in matching markets. *Management Science, 65*(12), 5603–5618.
Dubins, L. E., & Freedman, D. A. (1981). Machiavelli and the gale-shapley algorithm. *American Mathematical Monthly, 88*, 485–494.
Echenique, F., & Pereyra, J. S. (2016). Strategic complementarities and unraveling in matching markets. *Theoretical Economics, 11*(1), 1–39.
Ergin, H. I., & Sönmez, T. (2006). Games of school choice under the Boston mechanism. *Journal of Public Economics, 90*(1–2), 215–237.
Featherstone, C. (2020). Rank efficiency: Modeling a common policymaker objective. In *Working Paper*.
Featherstone, C. R., & Niederle, M. (2016). Boston versus deferred acceptance in an interim setting: An experimental investigation. *Games and Economic Behavior, 100*, 353–375.
Fragiadakis, D. E., & Troyan, P. (2019). Designing mechanisms to focalize welfare-improving strategies. *Games and Economic Behavior, 114*, 232–252.
Fréchette, G. R., Roth, A. E., & Ünver, M. U. (2007). Unraveling yields inefficient matchings: Evidence from post-season college football bowls. *The Rand Journal of Economics, 38*(4), 967–982.
Gale, D., & Shapley, L. S. (1962). College admissions and the stability of marriage. *American Mathematical Monthly, 69*, 9–15.
Ha, W., Kang, L., & Song, Y. (2020). College matching mechanisms and matching stability: Evidence from a natural experiment in China. *Journal of Economic Behavior & Organization, 175*, 206–226.
Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review, 92*(5), 1644–1655.
Kagel, J. H., & Roth, A. E. (2000). The dynamics of reorganization in matching markets: A laboratory experiment motivated by a natural experiment. *Quarterly Journal of Economics, 115*(1), 201–235.
Kesten, O. (2012). On two kinds of manipulation for school choice problems. *Economic Theory, 51*(3), 677–693.
Lien, J. W., Zheng, J., & Zhong, X. (2016). Preference submission timing in school choice matching: Testing fairness and efficiency in the laboratory. *Experimental Economics, 19*(1), 116–150.
Lien, J. W., Zheng, J., & Zhong, X. (2017). Ex-ante fairness in the Boston and serial dictatorship mechanisms under pre-exam and post-exam preference submission. *Games and Economic Behavior, 101*, 98–120.
Mongell, S., & Roth, A. E. (1991). Sorority rush as a two-sided matching mechanism. *American Economic Review, 81*, 441–464.

Niederle, M., & Roth, A. E. (2003). Unraveling reduces mobility in a labor market: Gastroenterology with and without a centralized match. *Journal of Political Economy, 111*(6), 1342–1352.
Pan, S. (2019). The instability of matching with overconfident agents. *Games and Economic Behavior, 113*, 396–415.
Roth, A. E. (1982). The economics of matching: Stability and incentives. *Mathematics of Operations Research, 7*, 617–628.
Roth, A. E. (1984). The evolution of labor market for medical interns and residents: A case study in game theory. *Journal of Political Economy, 92*, 991–1016.
Roth, A. E. (1985). The college admissions problem is not equivalent to the marriage problem. *Journal of Economic Theory, 36*(2), 277–288.
Roth, A. E. (1986). On the allocation of residents to rural hospitals: A general property of two-sided matching markets. *Econometrica, 54*(2), 425–427.
Roth, A. E. (1989). Two-sided matching with incomplete information about others' preferences. *Games and Economic Behavior, 1*(2), 191–209.
Roth, A. E. (1991). A natural experiment in the organization of entry-level labor markets: Regional markets for new physicians and surgeons in the United Kingdom. *American Economic Review, 81*, 414–440.
Roth, A. E., Sönmez, T., & Ünver, M. U. (2005). Pairwise kidney exchange. *Journal of Economic Theory, 125*(2), 151–188.
Roth, A. E., & Xing, X. (1994). Jumping the gun: Imperfections and institutions related to the timing of market transactions. *The American Economic Review*, 992–1044.
Sönmez, T. (1999). Can pre-arranged matches be avoided in two-sided matching markets? *Journal of Economic Theory, 86*(1), 148–156.
Sönmez, T., & Switzer, T. B. (2013). Matching with (branch-of-choice) contracts at the United States military academy. *Econometrica, 81*(2), 451–488.
Troyan, P. (2012). Comparing school choice mechanisms by interim and ex-ante welfare. *Games and Economic Behavior, 75*(2), 936–947.
Wang, T., & Zhou, C. (2020). High school admission reform in China: A welfare analysis. *Review of Economic Design, 24*(3), 215–269.
Wu, B., & Zhong, X. (2014). Matching mechanisms and matching quality: Evidence from a top university in China. *Games and Economic Behavior, 84*, 196–215.
Wu, B., & Zhong, X. (2020). Matching inequality and strategic behavior under the Boston mechanism: Evidence from china's college admissions. *Games and Economic Behavior, 123*, 1–21.