Games and Economic Behavior 115 (2019) 83–100

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Chinese college admissions and school choice reforms: An experimental study ☆

Yan Chen [a,c,*], Onur Kesten [b]

[a] School of Information, University of Michigan, 105 South State Street, Ann Arbor, MI 48109-2112, United States of America
[b] Tepper School of Business, Carnegie Mellon University, 5000 Forbes Avenue, PA 15213, United States of America
[c] Department of Economics, School of Economics and Management, Tsinghua University, Beijing, 100084, China

ARTICLE INFO

Article history:
Received 8 April 2016
Available online 19 February 2019

JEL classification:
C78
C92
D47
D82

Keywords:
College admissions
School choice
Immediate acceptance mechanism
Parallel mechanism
Deferred acceptance
Experiment

ABSTRACT

Since 2001, many Chinese provinces have transitioned from a "sequential" to a "parallel" school choice or college admissions mechanism. Inspired by this natural experiment, we evaluate the sequential (immediate acceptance, IA), parallel (PA), and deferred acceptance (DA) mechanisms in the laboratory. We find that participants are most likely to reveal their preferences truthfully under DA, followed by PA and then DA. While stability comparisons also follow the same order, efficiency comparisons vary across environments. Regardless of the metrics, the performance of PA is robustly sandwiched between IA and DA. Furthermore, 53% of our subjects adopt an insurance strategy under PA, making them at least as well off as what they could guarantee themselves under IA. These results help explain the recent reforms in Chinese school choice and college admissions.

1. Introduction

Two canonical mechanisms in the school choice and college admissions context are the Boston immediate acceptance mechanism (IA) and the student-proposing deferred acceptance (DA) mechanism (Gale and Shapley, 1962). While IA has been widely used in practice, game theoretic analysis (Abdulkadiroğlu and Sönmez, 2003; Ergin and Sönmez, 2006) and experimental evidence (Chen and Sönmez, 2006) indicate that it is vulnerable to strategic manipulation and thus might not

☆ This paper is the second part of a paper formerly entitled "Chinese College Admissions and School Choice Reforms: Theory and Experiments." We thank Susan Athey, Dirk Bergemann, Caterina Calsamiglia, Yeon-Koo Che, Isa Hafalir, Rustam Hakimov, Fuhito Kojima, Scott Kominers, Erin Krupka, Morimitsu Kurino, John Ledyard, Antonio Miralles, Herve Moulin, Parag Pathak, Jim Peck, Paul Resnick, Al Roth, Rahul Sami, Tayfun Sönmez, Guofu Tan, Utku Unver, Xiaohan Zhong and seminar participants at Arizona, Autonoma de Barcelona, Bilkent, Carnegie Mellon, Columbia, Florida State, Kadir Has, Johns Hopkins, Michigan, Microsoft Research, Rice, Rochester, Sabanci, Shanghai Jiao Tong, Stanford, Tsinghua, UCLA, UC-Santa Barbara, USC, UT-Dallas, UECE Lisbon Meetings (2010), the 2011 AMMA, Decentralization, EBES, Stony Brook, WZB, and NBER Market Design Working Group Meeting for helpful discussions and comments, Ming Jiang, Yingzhi Liang, Malvika Deshmukh, Tyler Fisher, Robert Ketcham, Tracy Liu, Kai Ou, Ben Spulber and Chao Tang for excellent research assistance. We are grateful to two anonymous referees and the Advisory Editor for their constructive comments which helped improve the paper. Financial support from the National Science Foundation through grants no. SES-0720943 and 0962492 is gratefully acknowledged.

* Corresponding author.
E-mail addresses: yanchen@umich.edu (Y. Chen), okesten@andrew.cmu.edu (O. Kesten).

https://doi.org/10.1016/j.geb.2019.02.003
0899-8256/

result in socially desirable assignments. As a result, it has been replaced by versions of DA in New York City (Abdulkadiroğlu et al., 2005b) and Boston (Abdulkadiroğlu et al., 2005a).

Despite serious concerns regarding potential manipulation, IA remains a widely-used assignment tool and is the subject of a strand of literature that revisits the mechanism in various settings for both school choice[1] and college admissions.[2] In this paper, we strive to shed light on the sources of these different points of view regarding IA and its comparison to DA in light of the recent school choice and college admissions reforms in China where IA used to be the predominant assignment mechanism until 2001. A mechanism that we believe could provide key insights is the *parallel mechanism* (PA) pioneered in Hunan for college admissions in 2001, in Shanghai for school choice in 2003, and later adopted by most of the provinces in China (Chen and Kesten, 2017).

Since 1952, China has implemented centralized college admissions. In the recent years 9-10 million high school seniors annually compete for 6-7 million seats at various universities. The Chinese college admissions (CCA) mechanisms are centralized matching processes via standardized tests called *gaokao*, with each province implementing an independent matching process. These matching mechanisms fall into two classes: sequential and parallel. The sequential mechanism is a priority matching mechanism, executed sequentially across tiers in decreasing prestige. In the sequential mechanism, each college belongs to a tier. Within each tier, IA is used. When assignments in the first tier are finalized, the assignment process in the second tier starts, and so on. A common complaint about the sequential mechanism is that high-scoring students often remained unassigned or under-matched due to poor strategizing in the ranking of colleges (Nie, 2007).

To mitigate this problem, the parallel mechanism has been adopted by a majority of the provinces as an alternative to the sequential mechanism. In the parallel mechanism, students are provided with choice-bands in which they can place several "parallel" schools. For example, a student's first choice-band can contain three schools, A, B, and C (in decreasing desirability). Schools process student applications by choice-bands, where no student loses his score advantage for any school he lists within the same choice-band until he is rejected from all schools in that choice-band. In China, this mechanism is widely believed to improve allocation outcomes and has been adopted by many provinces. The intuitive appeal of PA over its predecessor, IA, comes from the fact that it allows students to entertain both risky and safe options in their preferences at the same time. For example, in the example given (with a band size of three), a students is given "three shots", one of which can be used for a more desirable but risky option and another for a less desirable but safe option. In other words, PA allows students to retain their would-be assignment under IA as "insurance" options while keeping more preferred options within reach.

Concurrent with Chinese college admissions reforms, the transition from IA to PA also happened in the school choice context. To equalize access to school resources across students of different socioeconomic backgrounds, the Chinese government abandoned the previous merit-based middle school admissions mechanism in 1998, and replaced it with an open enrollment school choice mechanism where parents rank schools and schools select students using IA (Lai et al., 2009). Since then, students applying for middle schools are prioritized on the basis of their residence, whereas those applying for high schools are prioritized based on their municipal-wide exam scores. Using public middle school admissions data from Beijing Eastern City District, He (2014) investigates parents' behavior and finds that parents are overcautious in that they play "safe" strategies too often. Combining survey data, middle school choice data and High School Entrance Exam test scores from Beijing, Lai et al. (2009) find that children of parents who made mistakes in middle school selection were admitted to lower quality schools and achieved lower test scores on the High School Entrance Exam three years later. To our knowledge, the adoption of PA for school choice was pioneered in Shanghai in 2003,[3] although we are not aware of any systematic empirical evaluation of PA in the school choice context.

To investigate the theoretical properties of PA, Chen and Kesten (2017) (hereafter shortened as CK) formulate a parametric family of *application-rejection* mechanisms where each member is characterized by some positive number $e \in \{1, 2, \ldots, \infty\}$ of parallel and periodic choice-band sizes that allow the application and rejection process to continue before assignments are made permanent. As parameter e increases, we go from IA ($e = 1$) to PA ($e \in [2, \infty)$), and from those to DA ($e = \infty$). They show that members of this family become "more manipulable" (in the sense of Pathak and Sönmez, 2013) as e decreases. Further, when $e' = ke$ for some integer k, any member of the family indexed by e is "more stable" than the member indexed by e', but not vice versa. Specifically, this implies that, under truth-telling, when IA (resp. PA) selects a stable matching, then PA (resp. DA) also does so, but the converse statement is not true.

As DA is the only member of the application-rejection family of mechanisms with a focal dominant strategy (while other members often have multiple equilibria) and the theoretical stability and welfare comparisons in CK assume truth-telling,

[1] In stylized models, Abdulkadiroğlu et al. (2011) and Miralles (2009) emphasize possible *ex ante* welfare advantages of IA relative to DA. Featherstone and Niederle (2016) confirm these predictions in the laboratory. Troyan (2012) shows that these findings are sensitive to the assumptions on the priority structure. Kojima and Utku Ünver (2010) offer axiomatic characterizations of IA, whereas Kesten (2011) shows that, contrary to DA, IA is immune to manipulation attempts by schools through concealing capacity. Dur (2018) and Dur et al. (2018) propose modifications to IA to reduce its gaming and increase its stability aspects.

[2] As we discuss below, IA used to be the only mechanism for college admissions in China. The current Chinese college admissions mechanisms not only differ in their matching algorithm but also in the timing of preference submission by students. Recent empirical and experimental studies such as Wu and Zhong (2014), Lien et al. (2016), and Jiang (2016) find that if students submit preferences before taking the exam, the measurement error in the exam can be corrected via IA, which leads to matchings that are stable with regard to students' aptitudes.

[3] This mechanism was adopted in Shanghai for high school admissions in 2003, http://edu.sina.com.cn/l/2003-05-15/42912.html, retrieved on September 30, 2017.

it is important to investigate the properties of these mechanisms in the laboratory where subjects face coordination issues similar to those in real-life school choice situations and need not even engage in equilibrium play.[4] Therefore, our experimental hypotheses are only partially informed by the theoretical analysis in CK and our experiment can help us uncover other relevant behavioral issues which have important policy implications in practice. To this end we evaluate three members of this family – IA, the simplest PA, and DA – in two environments in the laboratory. As such, our approach in this paper can also be seen as an investigation of three alternative mechanisms, as complementary to, rather than a test of CK.

In both environments, we find that subjects are most likely to reveal their preferences truthfully under DA, followed by PA and then IA. Despite the fact that most subjects manipulate their preferences under IA and (to a lesser extent) PA, the directional comparisons in CK still hold, i.e., DA achieves a significantly higher proportion of stable outcomes than PA, which is more stable than IA. However, the efficiency comparison is sensitive to the environment. Regardless of the evaluation metrics, the performance of PA is robustly sandwiched between IA and DA. Furthermore, we find that 53% of our subjects successfully adopt an insurance strategy under PA, making them at least as well off as what they could guarantee themselves under IA. These results help explain the recent reforms in Chinese school choice and college admissions.

To our knowledge, our paper presents the first experimental evaluation of the Chinese parallel mechanism relative to IA and DA, as well as equilibrium selection in school choice mechanisms. As our experiment is implemented in the complete information setting, where subjects know their priority before submitting their ranking of schools, the results apply to school choice settings that employ standardized tests rather than a setting that uses lotteries.

2. School choice problem

In this section, we set up the school choice mechanism and define the mechanisms. A school choice problem (Abdulkadiroğlu and Sönmez, 2003) is comprised of a number of students each of whom is to be assigned a seat at one of a number of schools. Further, each school has a maximum capacity, and the total number of seats in the schools is no less than the number of students. We denote the set of students by $I = \{i_1, i_2, \ldots, i_n\}$, where $n \geq 2$. A generic element in I is denoted by i. Likewise, we denote the set of schools by $S = \{s_1, s_2, \ldots, s_m\} \cup \{\emptyset\}$, where $m \geq 2$ and $\emptyset$ denotes a student's outside option, or the so-called null school. A generic element in S is denoted by s. Each school has a number of available seats. Let q_s be the number of available seats at school s, or the **quota** of s. For each school, there is a strict priority order of all students, and each student has strict preferences over all schools. The priority orders are determined according to state or local laws as well as certain criteria of school districts.

A **school choice problem** consists of a collection of priority orders and a preference profile. A **matching** μ is a list of assignments such that each student is assigned to one school and the number of students assigned to a particular school does not exceed the quota of that school. A matching μ is **non-wasteful** if no student prefers a school with unfilled quota to his assignment. A matching μ is **Pareto efficient** if there is no other matching which makes all students at least as well off and at least one student better off.

Using the definition by Balinski and Sönmez (1999), the **priority of student i for school s is violated** at a given matching μ (or alternatively, student i justifiably envies student j for school s) if i would rather be assigned to s to which some student j who has lower $s-$priority than i, is assigned. A matching is **stable** if it is non-wasteful and no student's priority for any school is violated.

A **school choice mechanism**, or simply a mechanism φ, is a systematic procedure that chooses a matching for each problem. A mechanism is Pareto efficient (stable) if it always selects Pareto efficient (stable) matchings. A mechanism φ is **strategy-proof** if it is a dominant strategy for each student to truthfully report his preferences.

We next describe a family of mechanisms that are central to our study. IA, which is referred as the *sequential* mechanism in the context of Chinese college admissions, was the prevalent college admissions mechanism in China in the 1980s and 1990s. By 2018, it is completely abandoned in favor of variants of PA (see the online appendix of CK for a historical account of the Chinese college admissions). We next provide a parametric algorithm to describe a general class of mechanisms that nest IA, PA and DA.

Given student preferences, school priorities, and school quotas, we begin by outlining a parametric *application-rejection algorithm* that indexes each member of the family by a periodic[5] choice-band size e. This choice-band size represents the number of choices the algorithm goes through when allocations are tentative before they become final. For example, if $e = 2$, the allocation is finalized every two choices. Specifically,

[4] Indeed, a long stream of lab experiments show that participant behavior in the lab can diverge from what is expected by equilibrium predictions (Crawford et al., 2013).

[5] In China, many provinces implement asymmetric versions of this algorithm where the size of the choice-band also varies across rounds. For the purpose of the present paper it suffices to restrict attention to the symmetric class which readily embeds the three mechanisms we test in the lab. See CK for a description and analysis of the larger asymmetric class of mechanisms. As of 2018, in Shanghai, Zhejiang and Shangdong, a student can pick any e choices from the full list of choices, whereas the other provinces still retain the tier structure, which means that applicants are allowed to choose only from a pre-specified subset of schools for each round. In our model and experiment, we do not impose the tier structure.

Round t ≥ 0**:**

- Each unassigned student from the previous round applies to his $te+1$-st choice school. Each school s considers its applicants. Those students with the highest $s-$priority are tentatively assigned to school s up to its quota. The rest of the applicants are rejected.

In general,

- Each rejected student, who is yet to apply to his $te+e$-th choice school, applies to his next choice. If a student has been rejected from all his first $te+e$ choices, then he remains unassigned in this round and does not make any applications until the next round. Each school s considers its applicants. Those students with the highest $s-$priority are tentatively assigned to school s up to its quota. The rest of the applicants are rejected.
- The round terminates whenever each student is either assigned to a school or is unassigned in this round, i.e., he has been rejected by all his first $te+e$ choice schools. At this point, all tentative assignments become final and the quota of each school is reduced by the number of students permanently assigned to it.

The algorithm terminates when each student has been assigned to a school. At this point, all the tentative assignments become final. The mechanism that chooses the outcome of the above algorithm for a given problem is called the *application-rejection mechanism* (e), denoted by φ^e. This family of mechanisms nests IA and DA as extreme cases, PA as intermediate cases (Chen and Kesten, 2017). Specifically, the *application-rejection mechanism* (e) is equivalent to IA when $e=1$, PA when $2 \leq e < \infty$, and DA when $e=\infty$.

Within the family of application-rejection mechanisms, i.e., $e \in \{1, 2, \ldots, \infty\}$, IA is the only Pareto efficient mechanism, whereas DA is the only stable and strategy-proof mechanism.

3. Experimental design

We design our experiment to compare the performance of IA ($e=1$), the simplest version of PA ($e=2$) and DA ($e=\infty$) motivated by the theoretical characterization of the family of application-rejection mechanisms in CK. We choose the complete information environment to test the theoretical predictions, especially those on Nash equilibrium outcomes, as the corresponding theoretical framework is in complete information. This implies that the participants observe both preferences and priorities of all other participants. In real life implementations of school choice mechanisms, students typically do not observe their position in a lottery when they submit their preference rankings. Therefore, our experimental implementation framework applies to school choice settings that employ standardized tests, such as those in Shanghai high schools and Chicago exam schools, rather than a setting that uses lotteries, such as those in New York or Boston. The assumption that students know the preferences of other students is unrealistic in most situations. However, in trading off internal versus external validity, we opt for the former.

A 3(mechanisms) $\times$ 2(environments) factorial design is implemented to evaluate the performance of the three mechanisms, {IA, PA, DA}, in two different environments, a simple 4-school environment and a more complex 6-school environment. We use a more general priority structure than that in Chinese college admissions, so that our results might be applicable in both the school choice and the college admissions contexts.[6]

3.1. The 4-school environment

The first environment, which we call the **4-school environment**, has four students, $i \in \{1, 2, 3, 4\}$, and four schools, $s \in \{a, b, c, d\}$. Each school has one slot, which is allocated to one participant. Schools and students are geographically distributed. There are school zones. A student who lives within the zone of her district school has higher priority with her district school, whereas those who live in other school zones have lower priority at this particular school.

The payoffs for each student are presented in Table 1. The square brackets, **[ ]**, indicate the resident of each school district, who has higher priority in that school than other applicants. Payoffs range from 16 points for a first-choice school to 5 points for a last-choice school. Each student resides in her second-choice school.

For each session in the 4-school environment, there are 12 participants of four different types. Participants are randomly assigned types at the beginning of each session. At the beginning of each period, they are randomly re-matched into groups of four, each of which contains one of each of the four different types. Four schools are available for each group. In each period, each participant ranks the schools. After all participants have submitted their rankings, the server allocates the schools in each group and informs each person of his school allocation and respective payoff. The experiment consists of 20 periods to facilitate learning.

The priority order for each school is separately determined as follows: (1) a student who lives within a school district has high priority at that school. (2) students who do not live within a school district has low priority at that school. The

[6] In a follow-up study, we test the same set of mechanisms in the college admissions context where colleges have identical priorities (Chen et al., 2015).

Table 1
Payoff table for the 4-school environment.

	a	b	c	d
Payoff to Type 1	[**11**]	7	5	16
Payoff to Type 2	5	[**11**]	7	16
Payoff to Type 3	7	16	[**11**]	5
Payoff to Type 4	5	16	7	[**11**]

Table 2
Priority position for each type among low priority students for the 4-school environment.

	Type 1	Type 2	Type 3	Type 4
Periods 1-5	1	2	3	4
Periods 6-10	2	3	4	1
Periods 11-15	3	4	1	2
Periods 16-20	4	1	2	3

priority among the low priority students is based on their respective position in a priority queue. Furthermore, we change the priority queue every five periods ("*block*") to investigate whether participant strategies are conditional on their priority. In the first five periods, a student's priority position is the same as her type number. In each subsequent block of five periods, her priority position increases by one per block. Specifically, the priority position for each type in each five-period block is tabulated in Table 2.

At the beginning of each session, the experimenter reads the experimental instructions (Online Appendix B) out loud, which explains the experimental environment and procedure. Therefore, the participants are informed of the preferences of all other participants and the priority orders of all schools before submitting a rank-ordered list in each round.

For each of the 4 different queues, we compute the Nash equilibrium outcomes under IA and PA as well as under DA. For all four blocks, IA has a unique Nash equilibrium outcome, where each student is assigned to her district school. This unique Nash equilibrium outcome of IA, $\mu^{C/S}$, is also the college/student-optimal matching, which is Pareto inefficient, with an aggregate payoff of 44. Since each student is assigned to her second choice, $\text{rank}_i = 2$, for all i, the sum of ranks of their final allocation is 8.

$$\mu^{C/S} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ a & b & c & d \end{pmatrix}$$

For all four blocks, the matching $\mu^{C/S}$ is also a Nash equilibrium outcome under PA and DA. However, PA and DA have exactly one more Nash equilibrium outcome for all four cases, which is the following Pareto efficient matching μ^*, with the sum of ranks of 6 and an aggregate payoff of 54:

$$\mu^* = \begin{pmatrix} 1 & 2 & 3 & 4 \\ a & d & c & b \end{pmatrix}.$$

The Nash equilibrium profile that sustains outcome μ^* is the following (asterisks are arbitrary): $P_1 = (a, *, *, *)$, $P_2 = (d, b, *, *)$, $P_3 = (c, *, *, *)$, and $P_4 = (b, d, *, *)$. This is an equilibrium profile regardless of the priority order.[7] Note that, in this equilibrium profile, types 1 and 3 misrepresent their first choices by reporting their district school as their first choices, while types 2 and 4 report their true top choices.[8]

We now analyze participant incentives to reveal their true preferences in this environment. While truth-telling is a weakly dominant strategy under PA and DA, it is not a Nash equilibrium under IA. To see why truth-telling is a weakly dominant strategy under PA, recall that every player's district school is her second choice. By playing the truth-telling strategy, a player will be allocated to her district school in the worst case, given that the second choice is not disadvantaged under PA. A player has no incentive to report her district school as her first choice, since she will lose her chance to get into her most preferred school by doing so. The third and fourth choices are irrelevant in this case. Therefore, truth-telling is a weakly dominant strategy under PA in the 4-school environment. Table 3 summarizes our analysis on truth-telling and Nash equilibrium outcomes.

[7] This is a Nash equilibrium because, for example, if student 1 (or 3) submits a profile where she lists school d (resp. b) as her first choice, then she may kick out student 2 (resp. 4) in the first step but 2 (resp. 4) would then apply to b (resp. d) and kick out 4 (resp. 2) who would in turn apply to d (resp. b) and kick out 1 (resp. 3). Hence student 1 (or 3), even though she may have higher priority than 2 (resp. 4), she cannot secure a seat at b (resp. d) under DA.

[8] Note that types 1 and 3's manipulation benefits types 2 and 4, thus it does not violate truth-telling as a weakly dominant strategy, since type 1 (resp. 3) is indifferent between truth-telling and lying. If type 1 (resp. 3) reverts to truth-telling, she will then cause a rejection chain which gives everyone their district school, including herself. Therefore, she is not better off by deviating from the efficient but unstable Nash equilibrium strategy.

Table 3
Truth-telling and Nash equilibrium outcomes in the 4-school environment.

	Truthful preference revelation			Nash equilibrium outcomes		
	IA	PA	DA	IA	PA	DA
Block 1 (periods 1-5)	not NE	dominant strategy	dominant strategy			
Block 2 (periods 6-10)	not NE	dominant strategy	dominant strategy	$\mu^{C/S}$	$\{\mu^{C/S}, \mu^*\}$	$\{\mu^{C/S}, \mu^*\}$
Block 3 (periods 11-15)	not NE	dominant strategy	dominant strategy			
Block 4 (periods 16-20)	not NE	dominant strategy	dominant strategy			

Table 4
Payoff table for the 6-school environment.

	a	b	c	d	e	f
Payoff to Type 1	[**9**]	16	11	13	7	5
Payoff to Type 2	16	[**11**]	5	13	9	7
Payoff to Type 3	9	16	[**7**]	11	5	13
Payoff to Type 4	16	7	9	[**13**]	5	11
Payoff to Type 5	16	13	11	7	[**9**]	5
Payoff to Type 6	16	13	11	5	7	[**9**]

3.2. The 6-school environment

While the 4-school environment is designed to compare the mechanisms in a simple context, we now test the mechanisms in a more complex environment where student preferences are generated by school proximity and quality.

In this **6-school environment**, each group consists of six students, $i \in \{1, 2, \ldots, 6\}$, and six schools $s \in \{a, b, \ldots, f\}$. Each school has one slot. Following Chen and Sönmez (2006), each student's ranking of the schools is generated by a utility function, which depends on school quality, school proximity and a random factor. The normalized payoff table is reported in Table 4.

For each session in the 6-school environment, we include 18 participants of **six different types**. Participants are randomly assigned types at the beginning of each session. The experiment consists of 30 periods, with random re-matching into three groups of six in each period. Again, we change the priority queue every five periods.

Compared with the 4-school environment, the 6-school environment has a much larger set of Nash equilibrium outcomes. We examine the 6 different priority queues and compute the Nash equilibrium outcomes under each mechanism. The list of Nash equilibrium outcomes and their Pareto dominance relationship for each block is included in Online Appendix A. In this environment, the college-optimal DA outcome μ^c is always dominated by other stable or unstable Nash equilibrium outcomes. In comparison, the student-optimal DA outcome μ^s never dominates (or is dominated by) any unstable Nash equilibrium outcome unless it overlaps with μ^c. In block 2, when there are three stable Nash equilibrium outcomes, the student-optimal stable outcome, μ^s, dominates the other two.

Lastly, we present the efficiency analysis for the 6-school environment. The allocations that maximizes the sum of payoffs are the following ones, each leading to the sum of ranks of 13 with an aggregate payoff of 78.

$$\mu_1^* = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ b & d & f & a & e & c \end{pmatrix} \text{ or } \mu_2^* = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ b & a & f & d & e & c \end{pmatrix}.$$

In comparison, the No Choice benchmark, where each student is assigned to her district school, generates the sum of ranks of 22 with an aggregate payoff of 58.

3.3. Experimental procedures

In each experimental session, each participant is randomly assigned an ID number and is seated in front of a terminal in the laboratory. The experimenter then reads the instructions aloud. Subjects have the opportunity to ask questions, which are answered in public. Subjects are then given 10 minutes to read the instructions at their own pace and to finish the review questions. After everyone finishes the review questions, the experimenter distributes the answers and goes over the answers in public. Afterwards, participants go through 20 (resp. 30) periods of a school choice experiment in the 4-school (resp. 6-school) environment. At the end of the experiment, each participant fills out a demographics and strategy survey on the computer. Each participant is paid in private at the end of the experiment. The experiment is programmed in z-Tree (Fischbacher, 2007).

Table 5 summarizes the features of the experimental sessions. For each mechanism in each environment, we conduct four independent sessions between May 2009 and April 2012 at the Behavioral and Experimental Economics Lab at the University of Michigan.[9] The subjects are students from the University of Michigan. This gives us a total of 24 independent sessions and

[9] All IA and DA sessions were conducted between May 2009 and July 2010. However, we found a z-Tree coding error for the IA_6 treatment during our data analysis. Thus, four additional sessions were conducted in July 2011 for this treatment, to replace the corresponding sessions. PA sessions were conducted in March and April 2012.

Table 5
Features of experimental sessions.

Treatment	Mechanism	Environment	# Subjects × # sessions	Total # of subjects
IA_4	IA	4-school	12×4	48
PA_4	PA	4-school	12×4	48
DA_4	DA	4-school	12×4	48
IA_6	IA	6-school	18×4	72
PA_6	PA	6-school	18×4	72
DA_6	DA	6-school	18×4	72

360 participants (354 unique subjects).[10] Each 4-school session consists of 20 periods. These sessions last approximately 60 minutes. In comparison, each 6-school session consists of 30 periods. These sessions last approximately 90 minutes. The first 20-30 minutes in each session are used for instructions. The conversion rate is $1 = 20 points for all treatments. Each subject also receives a participation fee of $5, and up to $3.5 for answering the Review Questions correctly. The average earning (including participation fee) is $19.08 for the 4-school treatments, and $25.42 for the 6-school treatments. Experimental instructions are included in Online Appendix B. The data are available from the authors upon request.

4. Experimental results

In examining our experimental results, we first explore individual behavior and equilibrium selection, and then report our aggregate performance measures of the three mechanisms. We also investigate the sensitivity of our results to environment changes.

In presenting the results, we introduce several shorthand notations. First, let $x > y$ denote that a measure under mechanism x is significantly greater than the corresponding measure under mechanism y ($p \leq 0.05$). Second, let $x \geq y$ denote that a measure under mechanism x is greater than the corresponding measure under mechanism y, but the comparison is weakly significant ($0.05 < p \leq 0.10$). Lastly, $x \sim y$ indicates that the performance of the two mechanisms are not different from each other.

4.1. Individual behavior

We present our analysis of individual behavior in the order of truth-telling, manipulability, district school bias, best response, and Nash equilibrium. We first re-state each relevant theoretical result from the prior literature, formulate its empirical implications as a hypothesis, and then test the hypothesis using our experimental data.

We first examine the extent to which individuals reveal their preferences truthfully under each mechanisms. Since truth-telling is a weakly dominant strategy under DA (resp. PA) in both (resp. 4-school) environments, and not a Nash equilibrium under IA or PA_6, we expect that participants will be more likely to reveal their preferences truthfully under the strategy-proof mechanisms compared to the non-strategy-proof mechanism. This leads to our first hypothesis.

Hypothesis 1 *(Truth-telling).* The proportion of truthful preference revelation follows: (a) DA > IA in both environments, (b) DA ∼ PA, and (c) PA > IA in the 4-school environment.

Fig. 1 presents the proportion of truth-telling in the 4- and 6-school environments under each mechanism. Note that, under IA, truthful preference revelation requires that the entire reported ranking is identical to a participant's true preference ranking.[11] However, under DA and PA_4, truthful preference revelation requires that the reported ranking be identical to the true preference ranking from the first choice through the participant's district school, while the remaining rankings, from the district school to the last choice, are irrelevant. Under PA_6, truth-telling requires truthful reporting of the entire list except for Type 4 whose district school is her true second choice. Consequently, for Type 4, we require that the subject reports the top-two choices truthfully, as what is reported below is irrelevant. In general, under PA ($e = 2$), we can have multiple truth-telling strategies only if the district school is either the true first best or true second best. Note that the requirement of truthful reporting is much more demanding for IA than for DA and PA, purely for a mechanical reason.

While DA has a robustly higher proportion of truth-telling than IA, we find that PA also performs better than IA, especially in the 4-school environment. Tables C.1 and C.2 in Online Appendix C report the proportion of truth-telling (left panel) and district school bias[12] (right panel) under each mechanism by block in the 4- and 6-school environments, respectively. The bottom panel of each table reports the session-level comparisons. Results are summarized below.

[10] Despite our explicit announcement in the advertisement that subjects should not participate in the experiment more than once and our screening before each session, six subjects participated twice. We compare their behavior between their inexperienced and experienced sessions and do not find significant difference.

[11] The only exception is when a participant's district school is her top choice. In this case, truthful preference revelation entails stating the top choice. However, by design, this case never arises in our experiment, as no one's district school is her first choice.

[12] District School Bias will be defined later in this section.

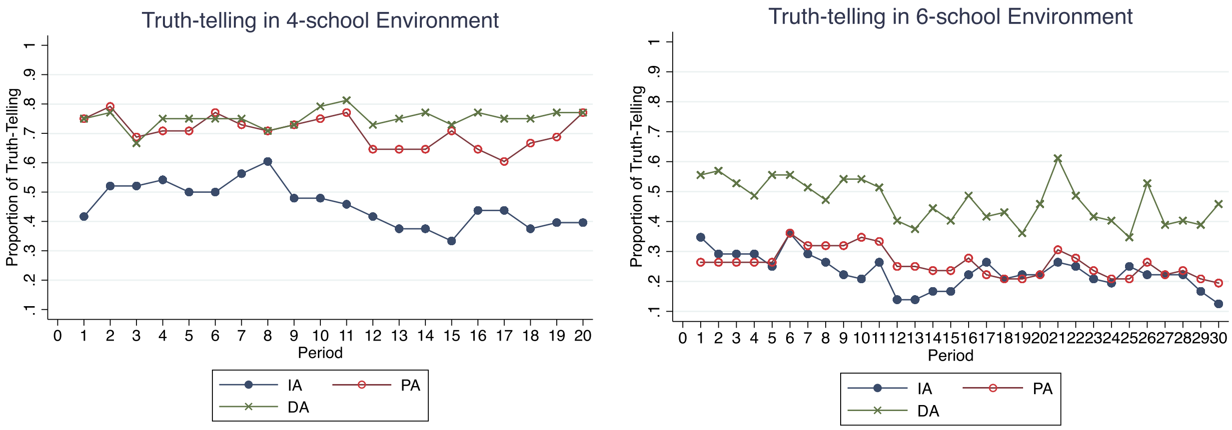

Fig. 1. Proportion of truth-telling in each environment.

Result 1 *(Truth-telling).* The proportion of truthful preference revelation follows the order: (a) DA > IA in both environments; (b) DA ∼ PA (resp. DA > PA) in the 4-school (resp. 6-school) environment; and (c) PA > IA in the 4-school environment.

SUPPORT: The left panels of Tables C.1 and C.2 present the proportion of truthful preference revelation by block and by session for the 4- and 6-school environments, respectively. P-values are computed from one-sided permutation tests, treating each session as an independent observation. ■

By Result 1, we reject the null in favor of Hypothesis 1(a) and 1(c). Furthermore, our result is consistent with Hypothesis 1(b). The result is similar for inexperienced participants (first period). While the ranking of truth-telling between IA and DA is consistent with Chen and Sönmez (2006), truth-telling under PA is reported for the first time.

While we do not observe 100% truth-telling under DA, it is significantly higher than IA in both environments and PA in the 6-school environment. Furthermore, we observe that the proportion of truth-telling under DA is significantly higher in the 4-school environment than in the 6-school environment ($p = 0.014$, one-sided permutation test). We interpret this as due to the relative simplicity of the environment.

The low rate of truth-telling under DA, especially in the 6-school environment, is related to the fact that DA is not obviously strategy-proof (Li, 2017). In strategically straightforward environments, Hassidim et al. (2016) find that a significant fraction of applicants ranked a non-funded position above a funded position in the same program in the Israeli psychology masters' match. They further find that weaker applicants in their field data set as well as those in Li's (2017) laboratory experiment (under random serial dictatorship) tend to misrepresent their preferences, likely due to misperceiving the rules of the mechanism. Using Australian college admissions data, Artemov et al. (2017) find that a non-negligible fraction of applicants adopt dominated strategies. However, the majority of these mistakes are payoff irrelevant. In fact, results suggest that DA would be largely stable even when some applicants do not report their preferences truthfully, which is consistent with our stability results in Fig. 6.[13]

Note that subjects are *not* told that truth-telling is a dominant strategy under DA in the experimental instructions (Online Appendix B). Following the convention in the experimental mechanism design literature, we describe each algorithm without prompting the subjects to behave in one way or another. Thus, results in this section summarize participant behavior without prompting from the experimenter. In practice, however, the market designer can educate the students when truth-telling is a dominant strategy. In fact, the Boston Public Schools, after switching to DA, advise the students to "list your school choices in your true order of preference" and that "there is no need to "strategize."[14] If parents follow the advice, we expect DA to achieve close to 100% truth-telling in practice, further enlarging the gap between DA and the other mechanisms reported in Result 1.

To investigate factors affecting truth-telling, we present six probit specifications in Table 6. The dependent variable is a dummy variable indicating whether a participant reveals her preferences truthfully. The independent variables include Priority Position (1 being the best, and 6 being the worst), the rank of a student's district school,[15] and a period variable to capture any effects of learning. In the 4-school environment (specifications 1-3), participants are 12.8 (resp. 7.1) percentage points less likely to tell the truth under IA (resp. PA) for a one-position worsening in the Priority Position, while such an effect is absent under DA, where truth-telling is a dominant strategy. We also observe a small but significant

[13] Stability under DA is similar in the 4- and 6-school environments, even though the proportion of truth-telling is significantly different.

[14] Source: http://www.bostonpublicschools.org/files/introbps_13_english.pdf, retrieved on December 12, 2013.

[15] This variable is only included for the 6-school environment, i.e., in specifications (4) to (6), as there is no variation in the position of the district school in the 4-school environment.

Table 6
Probit: truthful preference revelation.

Dependent variable: truth-telling						
Environments:	4-School environment			6-School environment		
Specifications:	(1)	(2)	(3)	(4)	(5)	(6)
Mechanisms:	IA	PA	DA	IA	PA	DA
Priority position	-0.128***	-0.071***	-0.013	-0.098***	-0.049***	-0.029***
	(0.027)	(0.010)	(0.018)	(0.010)	(0.009)	(0.005)
Rank of district school				-0.012	-0.143***	-0.112
				(0.025)	(0.043)	(0.073)
Period	-0.009**	-0.001	0.002	-0.004	-0.004***	-0.005**
	(0.004)	(0.004)	(0.003)	(0.003)	(0.000)	(0.002)
Log likelihood	-619.965	-519.034	-538.004	-985.820	-1,086.334	-1,429.058
Observations	960	960	960	2,160	2,160	2,160

Notes:
1. Robust standard errors are adjusted for clustering at the session level.
2. Coefficients are probability derivatives.
3. Significant at the: ** 5 percent level; *** 1 percent level.

effect of learning to manipulate under IA. In comparison, in the 6-school environment (specifications 4-6), we observe a similar Priority Position effect on truth-telling, but for all three mechanisms. The 2.9-percentage-point marginal effect of Priority Position on truth-telling under DA indicates that some participants might not understand the incentives in DA in the 6-school environment, consistent with the significantly lower level of truth-telling in this environment compared to the 4-school environment (Fig. 1). Again, we observe a small but significant effects of learning on preference manipulation under PA and DA mechanisms. We explore fatigue as an potential explanation for the lack of learning by considering only the first 20 periods, but obtain very similar results (Table C.3 in Online Appendix C). In specification (5), we observe that, when the ranking of a student's district school goes down by one position, a student's likelihood of truth-telling decreases by 14.3 percentage points under PA, whereas it has no significant effect on either IA or DA. Later in this section, we show that most of the district school bias under PA in the 6-school setting translates into an insurance strategy.

In the above analysis of truth-telling, as soon as a student inverts the relative ranking of two schools between the true and submitted preferences, she is marked untruthful. In Online Appendix C, we present our analysis of partial truth-telling, using the Kemeny distance to measure the extent of preference misrepresentation between two strategy profiles (Kemeny, 1959; Haeringer and Halaburda, 2016), and obtain rankings highly consistent with Result 1. Furthermore, one of the main results in Chen and Kesten (2017) is the manipulability comparison across this family of mechanisms. Using two metrics of manipulability, we obtain results which are largely consistent with the theoretical prediction (Online Appendix C.2).

Having examined manipulability across mechanisms, we now investigate patterns of manipulation. We first focus on District School Bias (DSB), a prevalent form of preference manipulation reported in previous experimental studies of IA (Chen and Sönmez, 2006; Calsamiglia et al., 2010; Klijn et al., 2012). Following (Chen and Sönmez, 2006), we define *District School Bias (DSB)* as a ranking pattern where a participant puts her district school into a higher-ranked position than that in the true preference order. Our next hypothesis is based on prior experimental findings as well as whether a mechanism is strategy-proof in a certain environment.

Hypothesis 2 *(District School Bias).* The proportion of DSB follows the order of IA $>$ PA $\sim$ DA in the 4-school environment, and IA $>$ PA $>$ DA in the 6-school environment.

Result 2 *(District School Bias).* The proportion of DSB follows the order of IA $>$ PA $\sim$ DA in the 4-school environment, and IA $\geq$ PA $>$ DA in the 6-school environment.

SUPPORT: The right panels of Tables C.1 and C.2 in Online Appendix C report the proportion of participants who exhibit DSB by block and over all blocks in the 4- and 6-school environment, respectively. The last column reports p-values from comparisons between mechanisms using one-sided permutation tests. ■

The comparison of DSB across mechanisms roughly mirrors that of truth-telling. Together, they account for roughly 80% of the strategies used by subjects. In what follows, we further highlight a particular form of DSB used in PA, identified by the public and formalized in our theory paper.

In China, PA is widely perceived by the public to improve allocation outcomes for students compared to IA. A main critique of IA is centered around the fact that the mechanism puts a lot of pressure on manipulation of first choices. PA alleviates this pressure. Theoretically, PA, in general (not true for the 4-school environment) does not make reporting first choice truthfully a dominant strategy. However, it frees up the pressure on the first choices, so under PA ranking your most preferred school as your top choice entails smaller potential cost relative to a similar decision under IA. Proposition 5 in

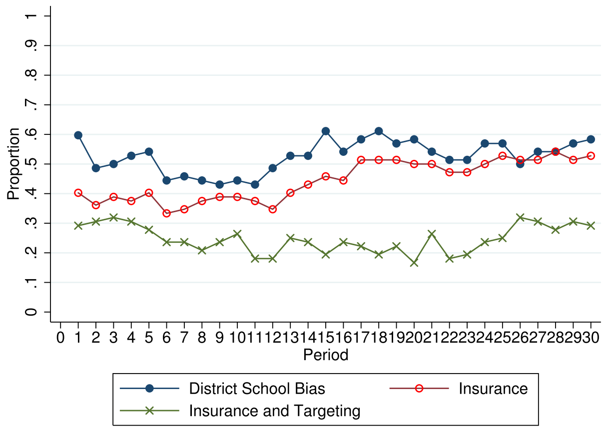

Fig. 2. Proportion of students adopting DSB, insurance, insurance and targeting strategies in the 6-school environment.

Chen and Kesten (2017) offers a formal argument that PA may indeed be more favorable for each student relative to the IA mechanism.

Proposition 5 *(Insurance under PA). Let μ be any equilibrium outcome under the IA mechanism. Under φ^e, if each student i lists $\mu(i)$ as one of his first e choices and also lists any schools he truly prefers over $\mu(i)$ as higher-ranked choices, then each student's parallel mechanism assignment is at least as good as his IA mechanism assignment.*

We now proceed to test the common perception among the general public in China that students can target their true first choices more often under PA than under IA. We now examine the likelihood that participants reveal their first choices truthfully under each mechanism.

Hypothesis 3 *(Truthful First Choice).* A higher proportion of reported first choices will be true first choices under PA than under IA.

Result 3 *(Truthful First Choice).* The proportion of truthful first choices under PA is significantly higher than that under IA in both environments.

SUPPORT: In the 4-school (6-school) environment, the proportion of truthful first choices is 78% (55%) under DA, 78% (48%) under PA, and 49% (37%) under IA. Using each session as an observation, one-sided permutation tests for pairwise comparisons of the proportion of truthful first choices yield PA > IA ($p = 0.014$ for the 4-school environment, $p = 0.029$ for the 6-school environment). ■

By Result 3, we reject the null in favor of Hypothesis 3 that PA generates a higher proportion of truthful first choices than IA. In particular, PA is virtually identical to DA in the proportion of truthful first choices in the 4-school environment. Regardless of the environment, participants are more likely to submit true first choices under PA than under IA.

One type of strategy implied by Proposition 5 is an insurance strategy. In our 6-school environment, students' district school position varies from the second (for Type 4) to the fifth position. To insure that she gets a school at least as good as her district school, she can put her district school as her second choice, and a more preferred school as her first choice. This is what we call an *insurance strategy*. Within this subset, if a student lists her most preferred school as her first choice and her district school as her second choice, we label it as the *insurance and targeting strategy*. Fig. 2 presents the proportion of students adopting DSB, insurance, as well as insurance and targeting strategies over time. By the last period, 58%, 53% and 29% of the subjects adopt DSB, insurance, and insurance and targeting strategies, respectively. That is, 91.4% (resp. 50%) of those who use DSB actually use the insurance (resp. insurance and targeting) strategy, confirming both the popular perception and the theoretical characterization of the insurance property of PA.

Having analyzed the individual truth-telling and manipulation strategies, we present our analysis of individual best response and Nash equilibrium strategies. We first investigate whether a student best responds to the other students' strategies in the same group. In the 4-school environment, truth-telling is a weakly dominant strategy under both PA and DA, whereas there is no dominant strategy under IA. Since a dominant strategy does not require that a subject has correct beliefs about others' strategies in order to best respond,[16] whereas she has to have the correct belief when a dominant

[16] The dominant strategy axiom was formulated as follows: "A fundamental, but generally unstated axiom of noncooperative behavior is that if an individual has a dominant strategy available, he will use it" (Groves and Ledyard, 1987, page 56).

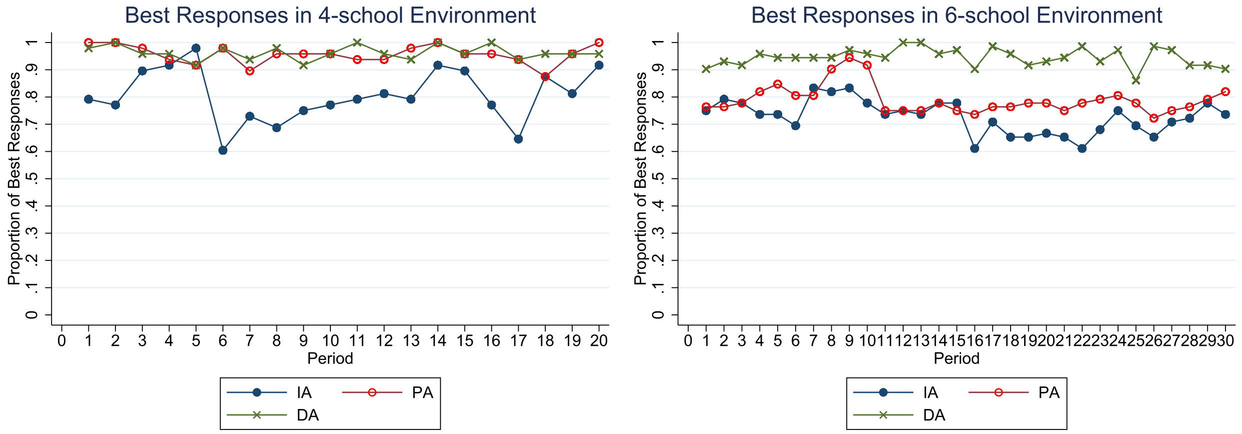

Fig. 3. Proportion of best responses in the 4- and 6-school environments.

strategy is absent, we expect a higher proportion of best responses under DA and PA than under IA. In comparison, we do not expect a significant difference in best response rate between DA and PA. By contrast, in the 6-school environment, while truth-telling is still the weakly dominant strategy under DA, there is no dominant strategy under either PA or IA. We formulate the following hypotheses.

Hypothesis 4 *(Best response).* The proportion of best responses follows the order of DA $\sim$ PA $>$ IA in the 4-school environment, DA $>$ PA $\sim$ IA in the 6-school environment.

Fig. 3 presents the proportion of best responses in the 4- and 6-school environments under each mechanism. We first observe that the proportion of best responses is high in the 4-school environment, greater than 80% under IA and 95% under PA or IA, but lower for IA and PA in the 6-school environment. The directional comparison is consistent with what we postulate in Hypothesis 4.

Result 4 *(Best response).* In each block, the proportion of best responses follows the order of DA$>$ IA, PA $>$ IA, and DA $\sim$ PA in the 4-school environment; DA $>$ IA, DA $>$ PA, PA $\sim$ IA in the 6-school environment.

SUPPORT: Tables C.10 and C.11 in Online Appendix C.3 present block- and session-level proportion of best responses (column 4), as well as p-values from permutation tests (column 5) in the 4- and 6-school environments, respectively. Results in Table C.10 indicate that block-level comparisons in the 4-school environment is consistent with the session-level comparisons: DA$>$ IA ($p < 0.05$ for all blocks), PA $>$ IA ($p < 0.05$ for all blocks), DA $\neq$ PA ($p > 0.10$ for all blocks). By comparison, results in Table C.11 indicates that block-level comparisons in the 6-school environment is consistent with the session level comparisons except in blocks 2 (PA $<$ DA, p-value $= 0.057$) and 4 (IA $\neq$PA, p-value $= 0.029$). ■

After our best response analysis, we decompose truth-telling into best-response and naïve truth-telling (see Table 7). The latter refers to the case when a subject reveals her preferences truthfully when it is not a best response, given the others' strategies in her group. Surprisingly, the majority of students who play truth-telling strategy best respond. When truth-telling is a dominant strategy, as in the 4-school environment under PA and in both environments under DA, 100% of the truth-telling are best-response truth-telling. When it is not a dominant strategy, as in IA in both environments and PA in the 6-school environment, between 15-20% of the truthful preference revelations are not best responses.

Moving from the analysis of individual best responses, we now present how often a group plays a Nash equilibrium profile, where every student in the group best responds to the others' strategies. We expect the Nash equilibrium play comparison resembles those of best responses.

Hypothesis 5 *(Nash equilibrium strategies).* In the 4-school environment, the proportion of groups which play Nash equilibrium follows the order: DA $\sim$ PA $>$ IA; In the 6-school environment, the proportion of groups which play Nash equilibrium follows the order: DA $>$ PA $\sim$ IA.

Fig. 4 presents the proportion of groups which play Nash equilibrium strategies in the 4- and 6-school environments under each mechanism, revealing the sharp drop in the proportion of Nash equilibrium play from the 4- to the 6-school environment for both IA ($0.446 \to 0.014$) and PA ($0.850 \to 0.114$).

Table 7
Decomposing truth-telling into best-response and naïve truth-telling.

(1) Mechanism	(2) Truth-telling	(3) Best response truth-telling	(4) Naïve truth-telling	(5) Naïve%
		4-school environment		
IA	0.456	0.379	0.077	0.169
PA	0.706	0.706	0.000	0.000
DA	0.751	0.751	0.000	0.000
		6-school environment		
IA	0.232	0.185	0.047	0.203
PA	0.258	0.217	0.041	0.159
DA	0.468	0.468	0.000	0.000

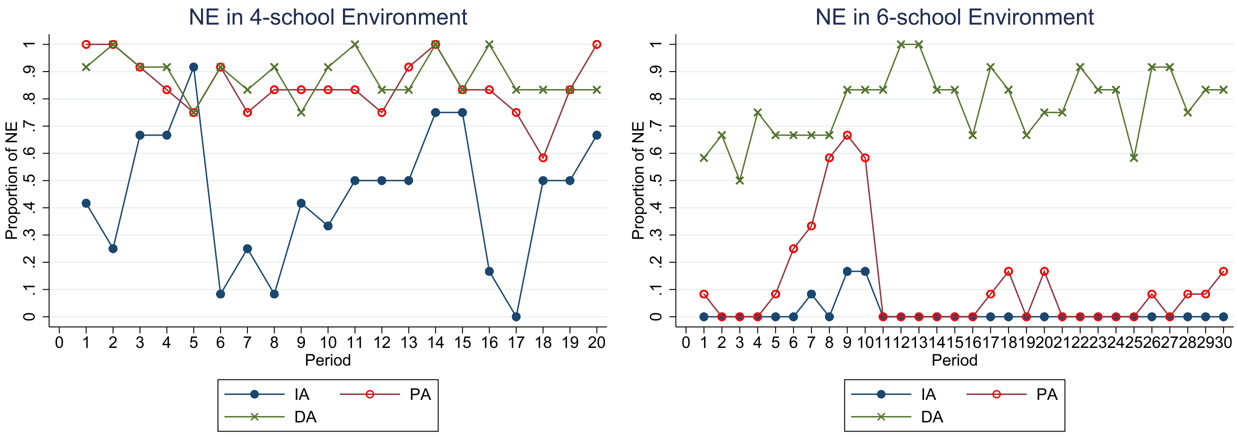

Fig. 4. Proportion of groups playing Nash equilibrium strategies in each environment.

Result 5 *(Nash equilibrium strategies).* In each block, the proportion of groups which play Nash equilibrium follows the order of DA $\sim$ PA $>$ IA; In the 6-school environment, the proportion of groups which play Nash equilibrium follows the order of DA $>$ PA $\sim$ IA.

SUPPORT: Tables C.10 and C.11 present block- and session-level proportion of groups playing Nash equilibrium strategies (column 6), as well as p-values from permutation tests (column 7) in the 4- and 6-school environments, respectively. Block-level comparisons are similar to session-level comparisons. Using each session as an independent observation, we find that (1) in the 4-school environment (Table C.10, column 7, bottom panel), DA $>$ IA ($p = 0.014$); PA $>$ IA ($p = 0.014$), and DA $\neq$ PA ($p = 0.800$); (2) in the 6-school environment (Table C.11, column 7, bottom panel), the proportion of groups playing Nash equilibrium follows DA $>$ IA ($p = 0.014$); DA $>$ PA ($p = 0.014$); PA $\neq$ IA ($p = 0.057$). ■

By Result 5, we reject the null in favor of Hypotheses 5. As with best responses, groups reaching a Nash equilibrium is robustly higher when there is a weakly dominant strategy. In the 6-school environment, one exception is in block 2, where the proportion of Nash equilibrium play under PA is not significantly lower than that under DA.

The proceeding analysis of best response and Nash equilibrium strategy analysis has implications for Nash equilibrium outcomes. Generically, there are multiple Nash equilibria in the application-rejection family of mechanisms. Thus, from both the theoretical and practical implementation perspectives, it is important to investigate which equilibrium outcomes are more likely to arise. To our knowledge, equilibrium selection in school choice mechanisms has not been studied before.

Our 4-school environment is particularly well suited to study equilibrium selection. Recall that in our 4-school environment, the student-optimal Nash equilibrium outcome, $\mu^{C/S}$, is the unique Nash equilibrium outcome under IA, while there are two Nash equilibrium outcomes under DA, $\mu^{C/S}$ and μ^{*}, where the latter Pareto dominates the former. Thus, it will be interesting to examine which of the two equilibrium outcomes arises more frequently under DA. While the Pareto criterion predicts that the Pareto optimal unstable Nash equilibrium should be selected, experimental results from secure implementation (Saijo et al., 2007) suggest that the dominant strategy equilibrium, when coinciding with the Nash equilibrium, is more likely to be chosen (Cason et al., 2006). This empirical finding is the basis for our next hypothesis.

Hypothesis 6 *(Equilibrium Selection).* Under DA, the stable Nash equilibrium outcome is more likely to arise compared to the unstable Nash equilibrium outcome.

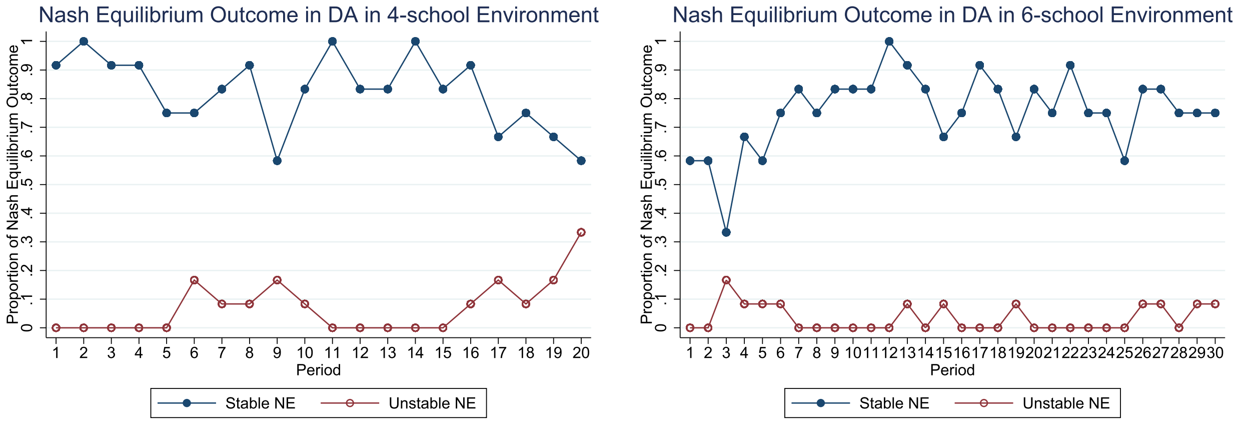

Fig. 5. Proportion of stable and unstable Nash equilibrium outcomes under the DA mechanism.

Table 8
Proportion of Nash equilibrium outcomes.

4-School	IA ($\mu^{C/S}$)	PA ($\mu^{C/S}$)	DA	DA ($\mu^{C/S}$)	DA (μ^*)	H_a	p-value
Session 1	0.683	0.933	0.967	0.950	0.017	IA $\neq$ PA	0.028
Session 2	0.600	0.817	0.850	0.717	0.133	IA $<$ DA	0.014
Session 3	0.600	0.867	0.817	0.800	0.017	PA $<$ DA	0.457
Session 4	0.533	0.633	0.950	0.833	0.117	DA(μ^*) $<$DA($\mu^{C/S}$)	0.063
6-School	IA	PA	DA	DA (Stable)	DA (Unstable)	H_a	p-value
Session 1	0.011	0.122	0.822	0.811	0.011	IA $\neq$ PA	0.028
Session 2	0.011	0.267	0.778	0.778	0.000	IA $<$ DA	0.014
Session 3	0.033	0.189	0.844	0.789	0.056	PA $<$ DA	0.014
Session 4	0.078	0.222	0.711	0.644	0.067	DA(unstable)$<$DA(stable)	0.063

Fig. 5 reports the proportion of the stable and unstable equilibrium outcomes over time under DA in the 4-school (left panel) and 6-school (right panel) environments, while Table 8 reports session-level statistics for each mechanism and pairwise comparisons between mechanisms and outcomes.

Result 6 *(Equilibrium Selection under DA).* Under DA, the proportion of the inefficient but stable Nash equilibrium outcome (82.5%) is weakly higher than that of the efficient but unstable Nash equilibrium outcome (8.9%) in the 4-school environment.

SUPPORT: In the 4-school environment, the average proportion of DA ($\mu^{C/S}$) is 82.5%, whereas the average proportion of DA (μ^*) is 8.9%. The last column in Table 8 presents the p-values for permutation tests comparing the proportion of equilibrium outcomes under different mechanisms. The null of equal proportion against the H_a of DA(μ^*) $<$ DA($\mu^{C/S}$) yields $p = 0.063$ (paired permutation test, one-sided). ■

By Result 6, we reject the null in favor of Hypothesis 6 at the 10% significance level. We conjecture that the stable Nash equilibrium outcome ($\mu^{C/S}$) is observed more often despite being Pareto dominated by μ^*, because the former requires truthful preference revelation, the weakly dominant strategy adopted by about 75% of the participants under DA, while the latter requires coordinated manipulation of top choices by players 1 and 3. However, we also note an increase of the unstable but efficient Nash equilibrium outcome, μ^*, in the last block in Fig. 5 (left panel), indicating that players 1 and 3 learn to coordinate their manipulation towards the end of the game. In fact, if we perform the same set of analysis on the last rounds in the 4-school environment, the comparison is no longer significant at the 10% level ($p = 0.125, 0.125, 0.375$ for the last 5, 3 and 1 rounds, respectively, paired permutation tests). This increase has direct implications for the efficiency comparisons in Result 8.

In comparison to the 4-school environment, the 6-school environment generates many Nash equilibrium outcomes. Because of this multitude of Nash equilibria, without strategy-proofness, on average, 3% and 20% of the outcomes are Nash equilibrium outcomes under IA and PA, respectively. In contrast, 79% of the outcomes under DA are Nash equilibrium outcomes. The proportion of this Nash equilibrium outcome follows DA $>$ IA ($p = 0.014$, one-sided permutation test), DA $>$ PA ($p = 0.014$, one-sided permutation test), and PA $>$ IA ($p = 0.014$, one-sided permutation test). If we break down the Nash equilibrium outcomes under the DA mechanism into stable and unstable equilibria, we again observe that the stable outcomes arise weakly more frequently than the unstable ones ($p = 0.063$, paired permutation test, one-sided). This effect continues to be significant at the 10% level if we only look at the last 5 or 3 rounds ($p = 0.062, 0.062$ for the last 5

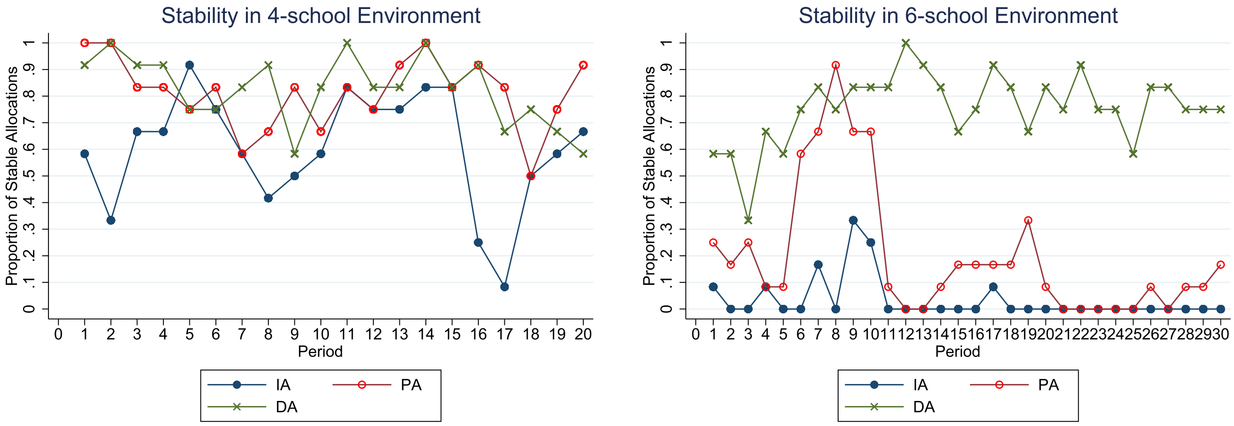

Fig. 6. Proportion of stable allocations in the 4- and 6-school environments.

Table 9
Stability: first block, last block and all periods.

	First block (periods 1-5)			Last block			All periods		
4-school	IA	PA	DA	IA	PA	DA	IA	PA	DA
Session 1	0.733	1.000	1.000	0.533	0.733	0.867	0.683	0.933	0.950
Session 2	0.533	0.867	0.733	0.333	0.867	0.733	0.600	0.817	0.717
Session 3	0.800	0.867	0.933	0.400	0.867	0.600	0.600	0.867	0.800
Session 4	0.467	0.667	0.933	0.400	0.667	0.667	0.533	0.633	0.833
6-school	IA	PA	DA	IA	PA	DA	IA	PA	DA
Session 1	0.000	0.067	0.800	0.000	0.000	0.867	0.011	0.122	0.811
Session 2	0.000	0.200	0.600	0.000	0.200	0.867	0.011	0.267	0.778
Session 3	0.000	0.067	0.333	0.000	0.133	0.933	0.033	0.189	0.789
Session 4	0.133	0.333	0.467	0.000	0.000	0.467	0.078	0.222	0.644

and 3 rounds, respectively, paired permutation tests), but ceases to be weakly significant if we only look at the last round ($p = 0.125$, paired permutation test).

In sum, Result 6 and our analysis of the 6-school data indicate that the stable Nash equilibrium outcome is more likely to arise than the unstable Nash equilibrium outcomes under the DA mechanism. To our knowledge, this is the first empirical result on equilibrium selection under DA.

4.2. Aggregate performance

Having presented the individual behavior and equilibrium outcomes, we now evaluate the aggregate performance of the mechanisms using two measures: stability and efficiency achieved under each mechanism.

First, we evaluate the stability achieved under each mechanism. A theoretical benchmark is Proposition 2 from Chen and Kesten (2017).[17] Proposition 2 implies that when IA (resp. PA) selects a stable matching for a problem, then PA (resp. DA) also selects a stable matching for the same problem, but the converse does not hold. While Proposition 2 assumes truth-telling under each mechanism, which is clearly not satisfied in the experiment, we nonetheless investigate the extent to which the theoretical ranking preserves when a large fraction of the students manipulate their preferences. We use three different metrics for our stability analysis, and report a binary measure in the main text. We define an allocation as stable if no one in a group of four (resp. six) students has justified envy.

Fig. 6 presents the proportion of stable allocations under each mechanism in the 4-school (left panel) and 6-school (right panel) environments. We summarize our stability analysis below.

Result 7 *(Stability).* DA and PA are each significantly more stable than IA in both environments. Furthermore, DA is significantly more stable than PA in the 6-school environment.

[17] We first re-state the first part of Proposition 2 from Chen and Kesten (2017):

Proposition 2 *(Stability). Let $e' > e$. If $e' = ke$ for some $k \in \mathbb{N} \cup \{\infty\}$, then $\varphi^{e'}$ is more stable than φ^{e}.*

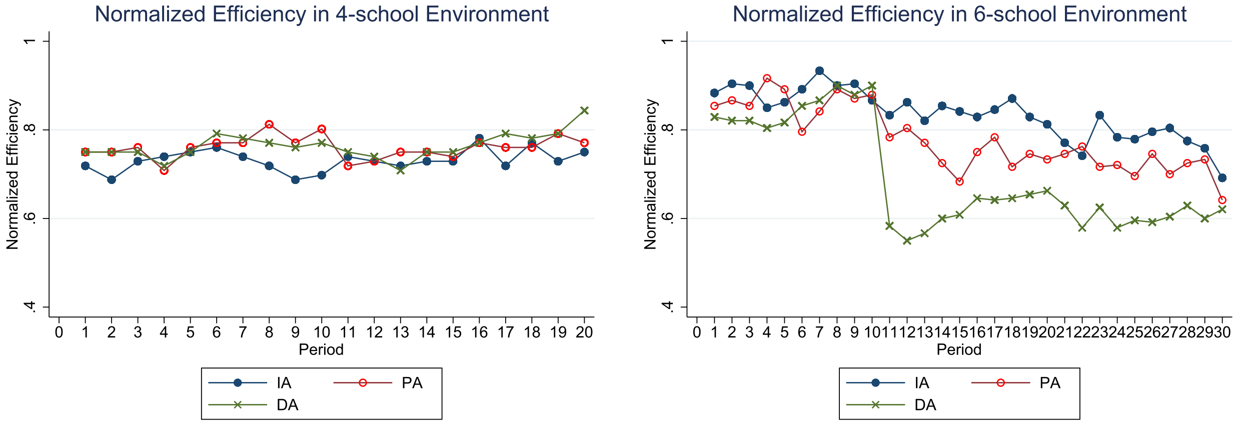

Fig. 7. Normalized efficiency in the 4- and 6-school environments.

SUPPORT: Table 9 reports the proportion of stable allocations among all allocations in the first and last block, and averaged over all periods in each session. Using one-sided permutation tests with each session as an observation, we find that (1) $DA_4 \geq PA_4$ ($p = 0.457$), $DA_4 > IA_4$ ($p = 0.014$), $PA_4 > IA_4$ ($p = 0.029$); (2) $DA_6 > IA_6$ ($p = 0.014$), $DA_6 > PA_6$ ($p = 0.014$), and $PA_6 > IA_6$ ($p = 0.014$). ■

Result 7 indicates that, in both environments, DA and PA each achieve a significantly higher proportion of stable allocations than IA. In the 6-school environment, DA also achieves a higher proportion of stable outcomes than PA. However, in the 4-school environment, the proportion of stable outcomes is indistinguishable between DA and PA, as both are strategy-proof. While our empirical stability ranking between DA and IA is consistent with Calsamiglia et al. (2010), the stability evaluation of PA is new.

In Online Appendix C.4, we report two more measures of stability - the proportion of students with justified envy, as well as the amount of highest justified envy a student suffers (as a small number of students are involved in multiple blocking pairs). Using three stability metrics, we find that DA is robustly stable across environments. While PA performs similarly as DA in the 4-school environment, its stability sandwiches between DA and IA in the 6-school environment. Lastly, IA is the least stable among the three mechanisms. The IA-DA gap is the largest using the binary metric, and smallest using the proportion of blocking pairs metric.

Lastly, we compare the efficiency of the mechanisms in each environment. In what follows, we present an efficiency measure using ordinal ranking of assignments.[18] We define a normalized efficiency measure as

$$\text{Normalized Efficiency} = \frac{\text{maximum group rank - actual group rank}}{\text{maximum group rank - minimum group rank}}, \qquad (1)$$

where *actual group rank* is the sum of the ranks of each assigned school in everyone's preference list. For example, if everyone is assigned to their district school in the 4-school environment, each assignment is ranked 2, as everyone's district school is their second most preferred school. The sum of the ranks is 8. Similarly, the *minimum group rank* is the sum of ranks for all group members for the Pareto efficient allocation(s), which equals 6 (resp. 13) for the 4-school (resp. 6-school) environment. Likewise, the *maximum group rank* is the sum of ranks for the worst allocation, which equals 14 (resp. 33) for the 4-school (resp. 6-school) environment. Because of this normalization, this measure always lies between zero and one, inclusive.

Fig. 7 presents the normalized efficiency under each mechanism in the 4-school and 6-school environments. Session-level normalized efficiency for the first and last blocks, as well as the average efficiency over all periods, is reported in Table 10.

Result 8 *(Efficiency).* In the 4-school environment, efficiency follows the order of $DA > IA$, and $DA \geq PA$. In the 6-school environment, efficiency follows the order of $IA > PA > DA$.

SUPPORT: Using one-sided permutation tests with each session as an observation, we find that
(1) First block: $IA_6 > DA_6$ ($p = 0.029$), $PA_6 > DA_6$ ($p = 0.029$), while none of the pairwise efficiency comparisons in the 4-school environment is significant.
(2) Last block: $DA_4 > IA_4$ ($p = 0.029$); $IA_6 > DA_6$ ($p = 0.014$); $PA_6 > DA_6$ ($p = 0.043$); $IA_6 \geq PA_6$ ($p = 0.057$).
(3) All periods: $DA_4 > IA_4$ ($p = 0.014$); $DA_4 \geq PA_4$ ($p = 0.071$); $IA_6 > PA_6$ ($p = 0.043$); $IA_6 > DA_6$ ($p = 0.014$); $PA_6 > DA_6$ ($p = 0.014$). ■

[18] For robustness check, we have also completed efficiency analysis based on the sum of payoffs, which yields similar results.

Table 10
Normalized efficiency: first block, last block and all periods.

	First block (periods 1-5)			Last block			All periods		
4-school	IA	PA	DA	IA	PA	DA	IA	PA	DA
Session 1	0.733	0.750	0.750	0.733	0.817	0.767	0.721	0.767	0.752
Session 2	0.742	0.758	0.733	0.758	0.733	0.808	0.744	0.752	0.777
Session 3	0.742	0.758	0.750	0.742	0.775	0.775	0.733	0.775	0.748
Session 4	0.683	0.717	0.742	0.767	0.758	0.833	0.727	0.746	0.777
6-school	IA	PA	DA	IA	PA	DA	IA	PA	DA
Session 1	0.870	0.887	0.800	0.773	0.753	0.567	0.849	0.805	0.676
Session 2	0.850	0.820	0.807	0.780	0.597	0.593	0.850	0.714	0.685
Session 3	0.910	0.907	0.850	0.740	0.710	0.560	0.810	0.801	0.679
Session 4	0.890	0.893	0.817	0.767	0.777	0.717	0.828	0.792	0.720

Result 8 indicates that DA is more efficient than IA in the 4-school environment. In the 6-school environment, since the proportion of Nash equilibrium play is low, theory has no clear prediction and the welfare comparison is likely to be sensitive to the setting. We find that a substantial fraction of subjects use effective heuristics, such as district school bias under IA (55%), insurance (53%), insurance and targeting (29%) under PA. These strategies are likely to bound losses and increase the likelihood of good payoffs. Finally, even though the proportion of equilibrium play is low, the proportion of best response is remarkably high (IA: 79%, PA: 73%). All of these might contribute to the high efficiency under IA and PA.

Under incomplete information, the efficiency comparison is ambiguous for settings closest to ours. For example, Ergin and Sönmez (2006) show that a Bayesian Nash equilibrium of IA need not be stable any more. Abdulkadiroğlu et al. (2011) show that IA could be more efficient in terms of interim payoff if students submit preferences before learning priority. This result is confirmed in an experiment by Featherstone and Niederle (2016). Lastly, Troyan (2012) introduces more priority structure and shows that neither DA nor IA dominates the other.

Result 8 also contributes to our understanding of the empirical performance of the school choice mechanisms. First, it indicates efficiency comparison is environment sensitive. While no single mechanism is more efficient in both environments, the parallel mechanism is never the worst. Second, while a first-period pairwise efficiency comparison is not significant in either environment, separation of performance occurs with learning, so that the last block ranking is significant. Our first period results are consistent with Calsamiglia et al. (2011). Our results point to the importance of allowing subjects to learn in school choice experiments. Lastly, our finding that DA is more efficient than IA in the last block is driven by the rise of the unstable but efficient Nash equilibrium outcome observed in Fig. 5 (left panel).

In sum, our experimental study has several new findings. First, we evaluate the performance of the simplest form of the parallel mechanism, and find that its manipulability, efficiency and stability measures are robustly sandwiched in between the IA and the DA mechanisms. Second, compared to the one-shot implementation of previous experiments on school choice except Featherstone and Niederle (2016),[19] our experimental design with repeated random re-matching enables us to compare the performance of the mechanisms with experienced participants. In doing so, we find that learning separates the performance of the mechanisms in terms of efficiency. Lastly, we report equilibrium selection under the DA mechanism for the first time, which reveals that stable Nash equilibrium outcomes are more likely to arise than the unstable ones even when the latter Pareto dominates the former.

5. Conclusions

While much of the debate on school choice exclusively focused on IA vs. DA mechanism comparisons, in this paper we provide a new dimension to this debate by bridging these mechanisms with the parallel mechanisms (PA) used for school choice and college admissions in China, and characterize them as members of a family of application-rejection mechanisms, with IA and DA being extremal members. Our theoretical analysis (Chen and Kesten, 2017) indicates a systematic change in the manipulability and stability properties of this family of mechanisms as one goes from one extreme member to the other.

To investigate behavioral responses to these mechanisms and to search for behavioral regularities where theory is silent, we conduct laboratory experiments in two environments differentiated by their complexity. In the laboratory, participants are most likely to reveal their preferences truthfully under DA, followed by PA and then IA. Furthermore, while DA is significantly more stable than PA, which is more stable than IA, efficiency comparisons vary across environments. Whereas theory is silent about equilibrium selection, we find that stable Nash equilibrium outcomes are more likely to arise than unstable ones. In another laboratory experiment designed to study the scale effect on the performance of matching mechanisms using the same environment as our simple 4-school environment in this paper and with complete information about payoffs

[19] Featherstone and Niederle (2016) investigate the performance of the IA and DA mechanisms under incomplete information, whereas we study the family of mechanisms under complete information. While their experiment is implemented under a random re-matching protocol, they do not explicitly analyze the effects of learning.

and priorities, we find that the mechanism rankings in our paper remains robust when the number of students per match increases from 4 to 40, and then to 4,000 (Chen et al., 2018).[20]

Our data also show that the proportion of groups playing Nash equilibrium under IA is low in the 4-school environment. It is even lower for both IA and PA in the more complex 6-school environment. The implication for theory and for future experiments is that Nash equilibrium might not be an appropriate solution concept in the school choice context. Researchers should explore alternative solution concepts, such as dominant strategy equilibrium or Bayesian Nash equilibria (see Abdulkadiroğlu et al., 2011 for IA and DA, and the appendix of Chen and Kesten, 2017 for PA under incomplete information).

Our study represents the first systematic experimental investigation of the class of Chinese parallel mechanisms. The analysis yields valuable insights which enable us to treat this class of mechanisms as a parametric family, and systematically study their properties and performance. More importantly, our results have policy implications for school choice and college admissions. As PA is less manipulable than IA, and its achieved efficiency is robustly sandwiched between the two extremes whose efficiency varies with the environment, it might be a less radical replacement for IA compared to DA. Our results indicate that a substantial fraction of subjects understand the insurance property of PA and effectively use it to reach for better schools without losing their safe schools.

In this first experimental study of PA relative to the two extreme mechanisms, we have investigated the performance of the simplest member ($e = 2$) of the PA family. Our results are encouraging for the market design literature: the observed behavior for PA is indeed sandwiched between those of IA and DA. This means that there are good reasons to believe that the way individuals respond to these mechanisms are in line with the theoretical predictions, albeit with substantial noise.

In practice, we have observed changes within the parallel family. For example, Hunan Province pioneered the parallel mechanism in 2001 which allowed three parallel choices per choice-band. Later, it switched to a different parallel mechanism allowing five parallel choices per choice-band in 2010. Using admissions data without the complete rank ordered lists, Wei (2015) find that, by 2013, the new parallel mechanism ($e = 5$) is significant more stable than the old parallel mechanism ($e = 3$). In future studies, it would be desirable to pick more members to investigate the performance of different PA mechanisms.

Around the same time as the Chinese college admissions reforms, a similar (but more drastic, in light of our analysis) transition took place in the United States in Boston and New York City in the context of school choice. Although economists were directly involved in the decision processes that lead to the reforms in these cities, we are not aware of any such involvement in the Chinese context. In China, college admissions is considered as "a battle that determines one's fate: one point [difference] in the exam can determine whether you go to heaven [i.e., universities] or hell [i.e., becoming a farmer]"(Yang, 2006). We suspect that such perception of extremely high stakes involved in this process and the search to reduce the element of gaming and risk-taking behavior might have been a strong driving factor in these reforms. Thus far the students and parents appear to have favorably responded to the reforms. As school choice and college admissions reforms continue in China and other parts of the world (Westkamp, 2013; Braun et al., 2014), experimental and empirical analyses of ongoing reforms offer insights which might affect education and labor market policies.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.geb.2019.02.003.

References

Abdulkadiroğlu, Atila, Sönmez, Tayfun, 2003. School choice: a mechanism design approach. Am. Econ. Rev. 93 (3), 729–747.

Abdulkadiroğlu, Atila, Pathak, Parag A., Roth, Alvin E., Sönmez, Tayfun, 2005a. The Boston public school match. Am. Econ. Rev. 95 (2), 368–371.

Abdulkadiroğlu, Atila, Pathak, Parag A., Roth, Alvin E., 2005b. The New York City high school match. Am. Econ. Rev. 95 (2), 364–367.

Abdulkadiroğlu, Atila, Che, Yeon-Koo, Yasuda, Yosuke, 2011. Resolving conflicting preferences in school choice: the "Boston Mechanism" reconsidered. Am. Econ. Rev. 101 (1), 399–410.

Artemov, Georgy, Che, Yeon-Koo, He, Yinghua, 2017. Strategic 'Mistakes': Implications for Market Design Research. Rice University Working paper.

Balinski, Michel L., Sönmez, Tayfun, 1999. A tale of two mechanisms: student placement. J. Econ. Theory 84, 73–94.

Braun, Sebastian, Dwenger, Nadja, Kübler, Dorothea, Westkamp, Alexander, 2014. Implementing quotas in university admissions: an experimental analysis. Games Econ. Behav. 85, 232–251.

Calsamiglia, Caterina, Haeringer, Guillaume, Klijn, Flip, 2010. Constrained school choice: an experimental study. Am. Econ. Rev. 100 (4), 1860–1874.

Calsamiglia, Caterina, Haeringer, Guillaume, Klijn, Flip, 2011. A comment on "School choice: an experimental study" [J. Econ. Theory 127 (1) (2006) 202–231]. J. Econ. Theory 146 (1), 392–396.

Cason, Timothy N., Saijo, Tatsuyoshi, Sjöström, Tomas, Yamato, Takehiko, 2006. Secure implementation experiments: do strategy-proof mechanisms really work? Games Econ. Behav. 57 (2), 206–235.

Chen, Yan, Kesten, Onur, 2017. Chinese college admissions and school choice reforms: a theoretical analysis. J. Polit. Econ. 125 (1), 99–139.

Chen, Yan, Sönmez, Tayfun, 2006. School choice: an experimental study. J. Econ. Theory 127, 202–231.

Chen, Yan, Jiang, Ming, Kesten, Onur, 2015. Chinese college admissions reforms: empirical and experimental evaluations. Manuscript.

Chen, Yan, Jiang, Ming, Kesten, Onur, Robin, Stéphane, Zhu, Min, 2018. Matching in the large: an experimental study. Games Econ. Behav. 110, 295–317.

[20] Chen et al. (2018) use all-human sessions for their 4- and 40-student per match treatments. When the number of students per match scales up to 4,000, each human subject is matched with 3,999 robots. These robots are programmed to play truthfully in one treatment ('truthful robot'), or to randomly draw a human strategy from a subject with similar priorities in the 40-subject per match treatment ('empirical robot').

Crawford, Vincent P., Costa-Gomes, Miguel A., Iriberri, Nagore, 2013. Structural models of nonequilibrium strategic thinking: theory, evidence, and applications. J. Econ. Lit. 51 (1), 5–62.
Dur, Umut Mert, 2018. The modified Boston mechanism. Math. Soc. Sci.
Dur, Umut, Hammond, Robert G., Morrill, Thayer, 2018. Identifying the harm of manipulable school-choice mechanisms. Am. Econ. J., Econ. Policy 10 (1), 187–213.
Ergin, Haluk I., Sönmez, Tayfun, 2006. Games of school choice under the Boston mechanism. J. Public Econ. 90 (1–2), 215–237.
Featherstone, Clayton R., Niederle, Muriel, 2016. Boston versus deferred acceptance in an interim setting: an experimental investigation. Games Econ. Behav. 100 (Suppl. C), 353–375.
Fischbacher, Urs, 2007. z-Tree: Zurich toolbox for ready-made economic experiment. Exp. Econ. 10 (2), 171–178.
Gale, David, Shapley, Lloyd S., 1962. College admissions and the stability of marriage. Am. Math. Mon. 69, 9–15.
Groves, Theodore, Ledyard, John O., 1987. Incentive compatibility since 1972. In: Groves, Theodore, Radner, Roy, Reiter, Stanley (Eds.), Information, Incentives and Economic Mechanisms: Essays in Honor of Leonid Hurwicz. University of Minnesota Press, Minneapolis, pp. 48–111.
Haeringer, Guillaume, Halaburda, Hanna, 2016. Monotone strategyproofness. Games Econ. Behav. 98 (Suppl. C), 68–77.
Hassidim, Avinatan, Romm, Assaf, Shorrer, Ran I., 2016. "Strategic" behavior in a strategy-proof environment. In: Proceedings of the 2016 ACM Conference on Economics and Computation. EC '16. ACM, New York, NY, USA, pp. 763–764.
He, Yinghua, 2014. Gaming the Boston School Choice Mechanism in Beijing. Rice University Working Paper.
Jiang, Ming, 2016. When Do Stable Matching Mechanisms Fail? The Role of Standardized Tests in College Admissions. University of Michigan Working Paper.
Kemeny, John G., 1959. Mathematics without numbers. Daedalus 88 (4), 577–591.
Kesten, Onur, 2011. On two kinds of manipulation for school choice problems. Econ. Theory, 1–17.
Klijn, Flip, Pais, Joana, Vorsatz, Marc, 2012. Preference intensities and risk aversion in school choice: a laboratory experiment. Exp. Econ., 1–22.
Kojima, Fuhito, Utku Ünver, M., 2010. The 'Boston' School Choice Mechanism. No. 729, Boston College Working Papers in Economics.
Lai, Fang, Sadoulet, Elisabeth, de Janvry, Alain, 2009. The adverse effects of parents' school selection errors on academic achievement: evidence from the Beijing open enrollment program. Econ. Educ. Rev. 28 (4), 485–496.
Li, Shengwu, 2017. Obviously strategy-proof mechanisms. Am. Econ. Rev. 107 (11), 3257–3287.
Lien, Jaimie W., Zheng, Jie, Zhong, Xiaohan, 2016. Preference submission timing in school choice matching: testing fairness and efficiency in the laboratory. Exp. Econ. 19 (1), 116–150.
Miralles, Antonio, 2009. School choice: the case for the Boston mechanism. In: Akan, Ozgur, et al. (Eds.), Auctions, Market Mechanisms and Their Applications. In: Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 14. Springer, Berlin, Heidelberg, pp. 58–60.
Nie, Haifeng, 2007. A game theoretic analysis of college admissions. China Econ. Quarterly 6 (3), 899–916.
Pathak, Parag A., Sönmez, Tayfun, 2013. School admissions reform in Chicago and England: comparing mechanisms by their vulnerability to manipulation. Am. Econ. Rev. 103 (1), 80–106.
Saijo, Tatsuyoshi, Sjöström, Tomas, Yamato, Takehiko, 2007. Secure implementation. Theor. Econ. 2 (3), 203–229.
Troyan, Peter, 2012. Comparing school choice mechanisms by interim and ex-ante welfare. Games Econ. Behav. 75 (2), 936–947.
Wei, Lijian, 2015. Chinese college admissions reform: natural experiments of matching. Wuhan University. Manuscript.
Westkamp, Alexander, 2013. An analysis of the German university admissions system. Econ. Theory 53 (3), 561–589.
Wu, Binzhen, Zhong, Xiaohan, 2014. Matching mechanisms and matching quality: evidence from a top university in China. Games Econ. Behav. 84, 196–215.
Yang, Xuewei, 2006. Commentaries on the History of the Chinese College Entrance Examination. Hubei People's Press, Wuhan.