An Empirical Evaluation of Chinese College Admissions Reforms Through A Natural Experiment*

Yan Chen Ming Jiang Onur Kesten

May 9, 2020

Abstract

College admissions policies have been widely debated around the world, affecting the education experiences and labor market outcomes for millions of students each year. In China alone, ten million high school seniors participate in the National College Entrance Exam to compete for seven million seats at various universities each year, making it the largest centralized matching market in the world. The last twenty years have witnessed radical reforms in the Chinese college admissions system, with many provinces moving from a sequential (immediate acceptance) system to some version of the parallel college admissions mechanism, which falls in between the immediate and deferred acceptance mechanisms. In this study, we use a natural experiment to evaluate the effectiveness of the sequential and parallel mechanisms in motivating student college ranking strategies and providing stable matching outcomes. Using a unique data set from a province that implemented a partial reform between 2008 and 2009, we find that students list more colleges in their rank-ordered lists, and more prestigious colleges as their top choices, after the province adopts the parallel mechanism in its Tier 1 college admissions process. These strategies in turn lead to greater stability in matching outcomes, consistent with our theoretical prediction that the parallel mechanism is less manipulable and more stable than the sequential mechanism.

Keywords: market design, college admissions, behavioral economics
JEL Classification: C78, C92, D47, D82

*We thank Yinghua He, Fuhito Kojima, Nancy Kotzian, Dorothea Kübler, Steve Leider, Sherry Xin Li, Tracy Xiao Liu, Alvin Roth, Utku Unver, Xiaohan Zhong and seminar participants at Stanford University, Tsinghua University, the University of Michigan, WZB and the Tsinghua BEAT Conference for helpful discussions and comments. Chen and Kesten gratefully acknowledge financial support from the National Science Foundation through grants no. SES-0720943 and SES-0962492. Chen: School of Information, University of Michigan, 105 South State Street, Ann Arbor, MI 48109-2112, USA. Email: yanchen@umich.edu. Department of Economics, School of Economics and Management, Tsinghua University, Beijing, 100084, China. Email: yanchen@umich.edu. Jiang: Antai College of Economics and Management, Shanghai Jiao Tong University, 1954 Huashan Road, Shanghai 200434, China. Email: mjiang@sjtu.edu.cn. Kesten: Tepper School of Business, Carnegie Mellon University, 5000 Forbes Avenue, PA 15213, USA. Email: okesten@andrew.cmu.edu.

1 Introduction

Since the 1990s, economic research has played an increasingly important role in the practical design of market institutions, including auctions for spectrums, electricity, and other commodities (Milgrom, 2004, 2019); tradable permit systems for pollution abatement and other environmental regulations (Ledyard and Szakaly-Moore, 1994); labor market clearinghouses (Roth, 1984, 1990; Kagel and Roth, 2000; Niederle and Roth, 2003); formal procedures for student assignments to public schools or colleges (Abdulkadiroğlu and Sönmez, 2003; Abdulkadiroğlu et al., 2005a,b); centralized systems for the allocation of organs (Roth et al., 2004); and other related matching and trading processes (Roth, 2015). In many of these cases, the insights drawn from theoretical, experimental, and empirical research have complemented each other in influencing market design choices.

Our study provides additional insight for the design of markets, obtained from a natural experiment to evaluate centralized matching procedures for student assignments to colleges. The college assignment process has a significant impact on the student educational experience as well as on broader labor market outcomes in countries that use a centralized college admissions system based on standardized test scores. These countries include China, Greece, Hungary, Russia, and Turkey.

Our study focuses on China in particular, where standardized test scores have been used since 1952 to match students to colleges via a centralized system. The National College Entrance Examination, also known as *gaokao*, forms the foundation of the Chinese college admissions system. Each year, roughly ten million high school seniors compete for seven million seats at various universities in China, making it the largest centralized matching market in the world (Chen and Kesten, 2017). Given the extent and importance of the Chinese admissions process, it is important to understand how the choice of mechanism impacts assignment outcomes.

The centralized college admissions problem (Balinski and Sönmez, 1999) has several unique properties compared to other matching problems such as school choice (Abdulkadiroğlu and Sönmez, 2003). One major differentiator is that students' priorities in college admissions are usually determined by their test scores on a standardized college entrance exam, rather than their place of residence, as in school choice problems. Therefore, college priorities are by and large identical across all colleges. Moreover, the prestige of a college is a major concern for virtually all students, leading to a near-universal preference for top universities with national prestige. This universal criterion implies that student preferences are often highly correlated. As a result, college admissions are typically much more competitive than student allocations to schools within a district. These two factors raise the stakes in the college admissions process and potentially affect how students strategize under different mechanisms.

In the past two decades, many Chinese provinces have moved from a "sequential" mechanism to various versions of a "parallel" mechanism in assigning students to universities. These universities are divided into tiers according to their level of prestige. The sequential mechanism is a priority matching mechanism (Roth, 1991) executed sequentially across tiers. Within each tier, the Immediate Acceptance (IA) mechanism is applied, e.g., once the assignments in the first tier are finalized, the assignment process in the second tier

starts, and so on. Despite its dominance in the admissions process until 2001, a common complaint about the sequential mechanism[1] is that many high-scoring students often remain unassigned or end up being under-matched due to poor strategizing in providing their preferred college rankings (Nie, 2007).

To combat this issue, Chinese provinces more recently have moved to some version of a parallel mechanism (PA), where students are provided with choice-bands in which they can list several "parallel" colleges in decreasing desirability. Under PA, student applications are processed by these choice-bands, wherein each student is guaranteed to retain his score advantage for any college she lists within the same choice-band. This mechanism is perceived to alleviate the pressure under IA by allowing students to "aim for" multiple colleges at the same time without the fear of losing their score advantage. For example, in Sichuan Province, where our dataset comes from, students can list up to five colleges within the same choice-band. Students can choose to allocate their choices across a mix of desirable-yet-risky and less-desirable-yet-safer options.

It is plausible to argue that the parallel mechanisms falls somewhere between the Immediate Acceptance and the Deferred Acceptance mechanism (Gale and Shapley, 1962). In a theoretical study of the Chinese college admissions reforms, Chen and Kesten (2017) (CK hereafter) consider a parametric family of *application-rejection* mechanisms where each member is indexed by some positive number $e \in \{1, 2, \ldots, \infty\}$ of periodic choice-band sizes that allow the application and rejection process to continue before assignments are made permanent. As parameter e increases, one goes from IA ($e = 1$) to PA ($e \in [2, \infty)$), and then to DA ($e = \infty$). CK show that members of this family become "more manipulable" (in the sense of Pathak and Sönmez (2013)) and "less stable" (in the sense of Kesten (2006)) as one moves away from DA. While multiple equilibria may arise under any member of the family, the important takeaway is that it is students' first e choices that matter. We use these theoretical insights as a partial basis for some of our hypotheses. Since the theoretical comparisons of IA and PA assume complete information and coordinated strategic play, it is important to test these predictions in the field to better gauge their policy implications.

Specifically, to complement and test the theoretical insights from CK, we study how different matching mechanisms in a centralized college admissions system affect students' preference ranking strategies and matching outcomes. Our study is based on a natural experiment that takes advantage of the move from IA to a parallel mechanism in Sichuan Province from 2008 to 2009. Since the year students participate in college admissions is mostly determined by birth, well before the change in mechanism, students in our study do not self-select into the different mechanisms. This timing eliminates any concern about selection bias in our study. Moreover, since the mechanism change affects only a portion of the students in our experiment (Tier 1 students), we are able to use a difference-in-differences approach. We find that the results from our experiment confirm some of the theoretical predictions from CK. Lastly, our data set is truly unique, as we have the complete rank-ordered list (ROL) from each applicant in addition to their matching outcomes,

[1] Such complaints are by now familiar from the school choice context where IA has come under extensive scrutiny due to its welfare and incentive shortcomings (Abdulkadiroğlu and Sönmez, 2003; Ergin and Sönmez, 2006; Chen and Sönmez, 2006; Pathak and Sönmez, 2008).

whereas other empirical studies of Chinese college admissions do not have the rank-ordered list for each applicant (Song et al., 2019; Wu and Zhong, 2014; Wei, 2009). The ROL data enable us to make more precise inference regarding students' strategic responses to the change in the matching mechanisms.

In particular, we find that students list more colleges in their ROL under PA relative to IA. Of the top-listed colleges, we observe a 5% increase in the preference rankings for the most prestigious colleges. Overall, our results show that the added "insurance" of being able to designate some safe options increases the stability of our matching outcomes.

2 Related Literature

Our study makes several important contributions to the literature on matching markets. Within this literature, a common way of testing matching mechanisms is via lab experiments. Doing so makes it possible to induce true preferences and thus accurately obtain various performance evaluations. Indeed, the school choice problem has been extensively studied using laboratory experiments that yield support for various mechanisms. For example, Chen and Sönmez (2006) find that DA performs well in terms of truthful preference revelation, while Pais and Pintér (2008) find that the Top Trading Cycles mechanism (TTC) is more efficient and less vulnerable to manipulation than either IA or DA in the school choice scenario; it is also less sensitive to the amount of information that participants hold. In experiments under the interim information condition, Featherstone and Niederle (2013) find that incomplete information on the students' side changes both mechanism efficiency and truthfulness, while Calsamiglia et al. (2010) find that constraining students' ability to reveal their preferences leads to greater manipulation and lower efficiency. We refer the reader to Hakimov and Kübler (2019) a recent review of the experimental matching literature.

Our paper contributes to the college admissions and broader matching literature by testing a common set of hypotheses using a natural experiment. The use of a field test provides higher external validity relative to laboratory experiments, since the latter is unable to capture the large scale and high stakes nature of the real-world college admissions process.

Field tests have been used to study the properties and performance of matching mechanisms. For example, Mongell and Roth (1991) study the "preferential bidding system" that matches students to sororities, and find that preference manipulation can prevent an unstable mechanism from unraveling. Braun et al. (2010) study the centralized college admissions in Germany, and find that high performing students who truth-tell due to a lack of understanding of the mechanism receive suboptimal placements. More recently, several studies have empirically examined the performance of matching mechanisms (Carvalho and Magnac, 2010; He, 2014; Calsamiglia et al., 2019; Agarwal and Somaini, 2018). These studies take a structural approach to the problem to uncover true preferences from reported ROLs when the mechanism is not strategy-proof. In a related study using school choice data from Beijing, He (2014) finds that teaching middle school parents to play best response under IA may yield better outcomes than switching to DA. Another strand of empirical literature, e.g., De Haan et al. (2015), Abdulkadiroğlu et al. (2017), Fack et al. (2019), and Kapor

et al. (2018), takes a more direct approach by using preference reports under strategy-proof mechanisms or surveys. In particular, Fack et al. (2019) provide theoretical and empirical evidence showing that, assuming stability of the matching provides rich identifying information, while being a weaker assumption on student behavior, compared to assuming that students truthfully rank schools when applying for admission. The latter is corroborated by an online experiment using medical students immediately after their participation in the medical residence match which features a strategy-proof market design (Rees-Jones and Skowronek, 2018).

Finally, in the Chinese school choice and college admissions context, the college admissions mechanisms not only differ in their algorithm but also in the timing of students' preference submissions. Wu and Zhong (2014) find that, under IA, better students are admitted to a top university when they submit their preferences before learning their test scores in the national college entrance exam. Using lab experiments, Lien et al. (2016) and Jiang (2016) argue that requiring preference submissions before students take the exam can help correct the observed exam measurement error under IA. However, Pan (2019) finds that pre-exam IA rewards overconfidence and creates more mismatches of students and schools. Comparing all three mechanisms in the Chinese school choice context in the laboratory, Chen and Kesten (2019) find that PA is less manipulable and more stable than IA. Our study contributes to this literature by examining how the adoption of PA impacts students' strategies and matching outcomes using a natural experiment.

3 Theory and Hypotheses

In this section, we introduce the college admissions problem, describe a family of mechanisms, and summarize the main theoretical results pertaining to this family. These theoretical results form the basis of our empirical evaluation.

We begin by defining the college admissions problem. Specifically, a *college admissions problem* (Balinski and Sönmez, 1999) is a tuple (S, C, P_s, P_C), consisting of: (1) a set of students $S = \{s_1, ..., s_n\}$; (2) a set of colleges $C = \{c_1, ..., c_m\}$; (3) a capacity vector $q = (q_{c_1}, ..., q_{c_m})$ where q_{c_i} is the capacity of college c_i; (4) a list of student preferences $P_S = (P_{s_1}, ..., P_{s_n})$ where P_{s_i} is the strict preference relation of student s_i over colleges including the no-college option (with an unlimited quota); and (5) a list of college preferences $P_C = (P_{c_1}, ..., P_{c_m})$ where P_{c_i} is the strict preference relation of college c_i over a set of students, determined by students' scores on the centralized college entrance exam. Therefore, $P_{c_i} = P_{c_j}$, $\forall i, j \in \{1, ..., m\}$. A matching μ is an allocation of college slots (and the no-college option) to students such that the number of students assigned to any college does not exceed its quota.

A matching μ is *nonwasteful* if no student prefers a college that has an unfilled quota. A matching is *stable* if it is nonwasteful and there is no student-college pair (c, s) such that student s prefers college c to the college she is assigned to, and college c prefers student s to at least one student who is assigned to it. A matching is *Pareto efficient* if there is no other matching that makes all students as well off and at least one student better off.

A college admissions mechanism, or simply a mechanism, is an algorithm that selects a matching for each problem. A mechanism is Pareto efficient (stable) if it always selects Pareto efficient (stable) matchings. A mechanism is *strategy-proof* if no student ever gains by misrepresenting his preferences.

Prior to 2001, the Immediate Acceptance, or sequential, mechanism was the prevalent college admissions mechanism in China. However, after 2001, a number of provinces began to adopt various versions of the parallel mechanism. By 2018, variants of PA had been adopted in all provinces. We next discuss an algorithm that describes a general family of mechanisms that nest IA, PA, and DA.

In the parametric *application-rejection algorithm* family, a member is indexed by a periodic choice-band size e that represents the number of choices the algorithm goes through when allocations are tentative before they become final.[2]

Round t $\geq$0:

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

- The round terminates whenever each student is either assigned to a college (including the no-college option) or is unassigned in this round, i.e., he has been rejected from all his first $te+e$ choices. At this point, all tentative assignments become final and the quota of each college is reduced by the number of students permanently assigned to the college.

The algorithm terminates when each student has been assigned to a college or has received the no-college option. At this point, all the tentative assignments become final. This family of mechanisms nests IA and DA as extreme cases, PA as intermediate cases (Chen and Kesten, 2017). Specifically, one obtains IA when $e=1$, PA when $2 \leq e < \infty$, and DA when $e=\infty$. In this family, IA is the only Pareto efficient mechanism, whereas DA is the only stable or strategy-proof mechanism. In Sichuan Province, which we study, $e=5$.

[2]Several provinces use asymmetric versions of this algorithm where the size of the choice-band also varies across rounds. See CK for further explanation of these variations as well as a historical account of the Chinese college admissions process.

In this family, IA is the only Pareto efficient mechanism, whereas DA is the only stable or strategy-proof mechanism.

In their theoretical study, CK find that a move from one extreme member to the other yields a trade-off in terms of strategic immunity and stability. At the individual strategy level, they show that, whenever any given member can be manipulated by a student, any member with a smaller e number can also be manipulated but not vice versa (Theorems 1 & 3). This implies that the PA mechanism used in Sichuan (where $e = 5$) is less manipulable than the previously-used IA mechanism. This leads to our first hypothesis:

Hypothesis 1 (Manipulability)**.** Students will manipulate their preferences less under PA compared to IA.

In our field setting, although true preferences are not directly observable, we can infer preference manipulation through a number of patterns, such as dropping some colleges from true rankings or listing a safe college as one's top choice, where a safe option may be a less prestigious college or a local college which allocates more spaces for local students.

Continuing with the theoretical predictions of CK, they suggest that students under PA are able to list their equilibrium assignments under IA as a safety option while also listing their more desirable options higher up in their preference list, which yields an outcome at least as good as that under IA (Proposition 5 in Chen and Kesten (2017)). While the top choice is the most critical decision under IA, the first e choices are of utmost importance when determining final assignments under PA. This leads to our next hypothesis:

Hypothesis 2 (Insurance)**.** Students will list more prestigious/more preferred colleges as their first choices under PA, compared to the IA mechanism.

In terms of choice accommodation, CK show that the IA mechanism is more generous in allocating students to their first choice than PA. This leads to the following hypothesis:

Hypothesis 3 (Choice Accommodation)**.** IA will assign a higher number of students to their reported first choices than will PA.

In terms of stability, CK show theoretically that the PAs in China are more stable than the IAs they replace (Theorems 2 and 4). This leads to our final hypothesis:

Hypothesis 4 (Stability)**.** PA will be more stable than IA.

Finally, we note that CK find no clear dominance of DA over PA, or PA over IA, due to the multiplicity of equilibria, even though (Ergin and Sönmez, 2006) find that the dominant strategy equilibrium outcome of DA Pareto dominates any equilibrium outcome of IA. Based on the predictions of CK, we are agnostic with regard to the efficiency comparison of the two mechanisms in our study.

4 Data and Empirical Methods

Our dataset consists of the college admissions data of a County in the Sichuan Province in southwestern China for the years 2008 and 2009.[3] The County had a population of 1.47 million in 2008 and 2009. It is 87.0% rural in 2008, and 86.6% in 2009,[4] with a per capita GDP of USD 994.3 in 2008 and 1117.1 in 2009,[5] below the national average of USD 3523.8 and 3827.5, respectively.[6] For our study, we obtain the following student data: test score on the National College Entrance Exam, rank-ordered list of colleges, college admission outcome, and demographics. Compared to prior empirical studies of Chinese college admissions, our data set is unique in that we have each student's rank-ordered list.

Chinese colleges are categorized into tiers of decreasing prestige and quality. For example, Tier 1 colleges are generally considered better than Tier 2 colleges, etc. To determine college placement assignments, admissions mechanisms are executed sequentially across tiers. When assignments in the first tier are finalized, the assignment process in the second tier starts, and so on. Our dataset contains all students who participated in the Tier 1, Tier 2, and Tier 3 admissions process in 2008 and 2009.

For the period of our dataset, students first received their test scores and relative standings among all the students in the province and then completed their rank-ordered lists of colleges. The Provincial College Admissions Office determined whether a student was eligible to participate in the admissions of each tier by setting up an endogenously determined cutoff score, such that the number of students who were above the Tier 1 cutoff was approximately 120% of the total quota of all Tier 1 colleges; the number of students who were above the Tier 2 cutoff was approximately 120% of the total quota of all Tier 1 and Tier 2 colleges; etc.[7]

Additionally, there were two separate matching markets each year for the two academic tracks: humanities and social sciences (shortened as humanities henceforth), and science and engineering (shortened as science henceforth). Students select into one of the two tracks in their second year of high school, and subsequently prepare for and then take the corresponding set of exams. Likewise, each college has a separate quota for each of the two tracks.

Between the college entrance exams of 2008 and 2009, the government of Sichuan Province announced that it would change the college admissions mechanism from IA to PA for only its Tier 1 selection process. Since students participate in the college admissions process during their last year of high school, and the policy change was announced after the previous year's admission was complete, students were selected into different treatment groups by birth. Thus, this context allows us to use this policy change as a natural experiment to study the effects of different matching mechanisms on students' behavioral responses and college admissions outcomes.

[3] Our data came from Santai County in Sichuan Province. The college admissions officials in Santai have requested that the county remains anonymous in any publication.

[4] The national rural population was 53.01% and 51.66% for 2008 and 2009, respectively. Source: National Bureau of Statistics

[5] Source: County Bureau of Statistics.

[6] Source: National Bureau of Statistics.

[7] See the online appendix of Chen and Kesten (2017) for a detailed discussion of the Chinese college admissions process.

Even though students are randomly selected into the different years by birth, we consider the possibility that there may be other differences across the two years, such as students' overall preferences for humanities versus science programs, that may impact our results. To address this possibility, we exploit the fact that only the Tier 1 mechanism changed from 2008 to 2009 in Sichuan Province, whereas the Tier 2 admissions mechanism remained the same. Therefore, we estimate the following difference-in-differences model:

$$y_i = \beta_0 + \beta_1 \cdot Y2009_i + \beta_2 \cdot Tier1_i + \beta_3 \cdot (Y2009_i \cdot Tier1_i) + \gamma \cdot \boldsymbol{X}_i + \epsilon_i,$$

where y_i is the outcome variable, measuring strategies or matching outcomes for each student. $Y2009_i$ and $Tier1_i$ are dummy variables that equal one for Year 2009 and Tier 1, respectively, and zero otherwise. The vector, $\boldsymbol{X}_i$, contains students' individual characteristics, including gender, residential status (rural or urban), academic track (humanities or science), and rank by test scores.

Table 1: Summary statistics

	2008				2009			
	Total	Female	Rural	Science	Total	Female	Rural	Science
Participated in Tier 1 admission	620	32.3%	80.2%	81.8%	768	36.3%	79.7%	80.7%
Participated in Tier 2 admission	2443	40.4%	80.8%	70.9%	2735	42.7%	83.1%	73.8%
Participated in Tier 3 admission	688	40.4%	75.0%	50.7%	605	48.3%	73.2%	57.4%
Participated in Tiers 1 and 2	122	43.4%	81.1%	77.9%	135	46.7%	81.5%	77.0%
Participated in Tiers 1, 2 and 3	2	100.0%	50.0%	100.0%	3	100.0%	100.0%	33.3%
Submitted Tier 1 ROL	717	30.54%	79.9%	82.0%	849	35.7%	79.5%	80.1%
Submitted Tier 2 ROL	2967	38.5%	80.7%	72.8%	3343	41.3%	82.7%	74.6%
Submitted Tier 3 ROL	876	49.2	72.3%	52.6%	787	48.9%	72.7%	57.8%
Submitted Tiers 1 and 2 ROL	628	33.0%	80.6%	80.4%	723	37.2%	80.8%	78.2%
Submitted Tiers 1, 2 and 3 ROL	4	100.0%	25.0%	25.0%	7	71.4%	57.1%	28.6%

Table 1 presents the summary statistics for our dataset. These statistics shows that students in different academic tracks and from different demographic backgrounds are similarly distributed across both years and tiers. Note that some students who were eligible for but did not receive a Tier 1 admission placement subsequently participated in the Tier 2 process. They submitted ROLs for both Tiers 1 and 2.

5 Results

In this section, we first report our results regarding student strategies and then discuss our results regarding matching outcomes for the Tier 1 admissions process, using the Tier 2 process as our control. In the appendix, we use Tier 3 as the control condition as a robustness check.

Table 2 reports the summary statistics for the main outcome variables. At the individual strategy level, we investigate both the change in the number of colleges students rank (length of ROL), as well as the change in the prestige status of their top choice colleges. At the outcome level, we examine the proportion

Table 2: Summary statistics for outcome variables

	Tier 1		Tier 2		Tier 3	
	2008	2009	2008	2009	2008	2009
Length of rank-order list (1-5)	3.676	4.473	4.166	4.239	3.245	3.392
Normalized top choice college prestige (0-1)	0.519	0.458	0.403	0.368	0.458	0.406
First choice accommodation rate	0.740	0.501	0.673	0.672	0.638	0.536
Stability based on college prestige	0.109	0.086	0.134	0.117	0.170	0.189
Stability based on cutoff score	0.118	0.070	0.273	0.292	0.237	0.358
Stability based on score distance	0.494	0.374	0.521	0.508	0.328	0.374

of students admitted to their top choices, as well as the stability of the matching outcomes, using several measures to ensure robustness.

5.1 Student Preference Ranking Strategies

In the college admissions process, students within a given tier are asked to rank order anywhere from one to five colleges. Figure 1 reports the average length of the ROLs in 2008 and 2009 by tier. The red solid (green dashed) line refers to the length of the ROLs of Tier 1 (2) students. From Figure 1, we see an increase in the ROL length for Tier 1 students from 2008 to 2009 by approximately one more college, whereas the average ROL length for Tier 2 students remains the same across two years. Figure 2 presents the distribution of the length of ROLs and similarly depicts an increase in the number of colleges submitted by Tier 1 students.

Table 3 presents the results from three OLS specifications. The dependent variable is the length of students' submitted ROL. The independent variables (omitted) include Y2009 (Y2008), Tier1 (Tier2), Y2009 $\times$ Tier1, Science (Humanities), Rural (Urban), Female (Male), and Percentile Ranking. To determine a student's percentile ranking, we calculate rankings in each of the 8 markets based on student test scores on the respective national college entrance exams, as matching is carried out separately by year (2008/2009), tier (1/2), and academic track (humanities/science). To correct for different market sizes, we then normalize student rankings to their percentile rankings in their respective markets (the top ranked student in each market is 1 (100%), and the bottom ranked 0). This measure of student rankings as our independent variable is used in all subsequent regressions.

The results in Table 3 show a positive significant coefficient for our main treatment effect, Y2009 $\times$ Tier1, indicating that the change from IA to PA in Tier 1 admissions process in 2009 increases the average ROL length by 0.724 ($p < 0.01$). That is, a Tier 1 student lists approximately one more college in 2009. Since the Tier 2 ROL length remains stable across the two years, the change in the length of the Tier 1 ROL is likely due to the change in the matching mechanism. Continuing with Table 3, we see that the coefficient for the Tier 1 dummy is negative and significant, indicating that the average length of the Tier 1 ROL is shorter than the corresponding Tier 2 ROL in 2008. This finding may reflect the importance of a student's first choice under IA, whereas lower-ranked colleges, such as a student's fourth or fifth choice, are not that useful under IA. Under PA, however, students have an incentive to include a less-prestigious college as their

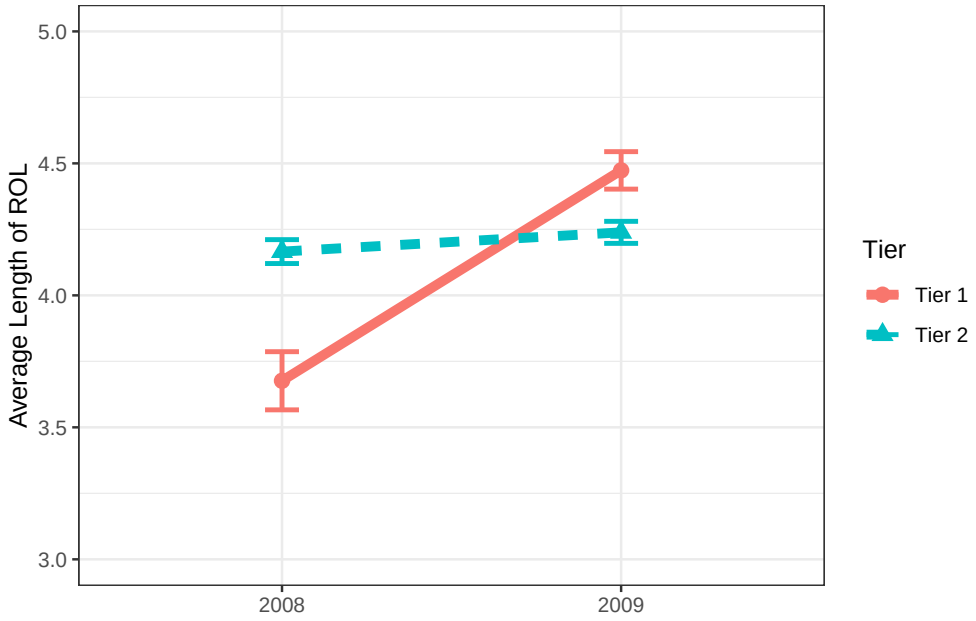


Figure 1: Average length of ROLs by year and tier; error bars indicate the 95% confidence interval of the mean.

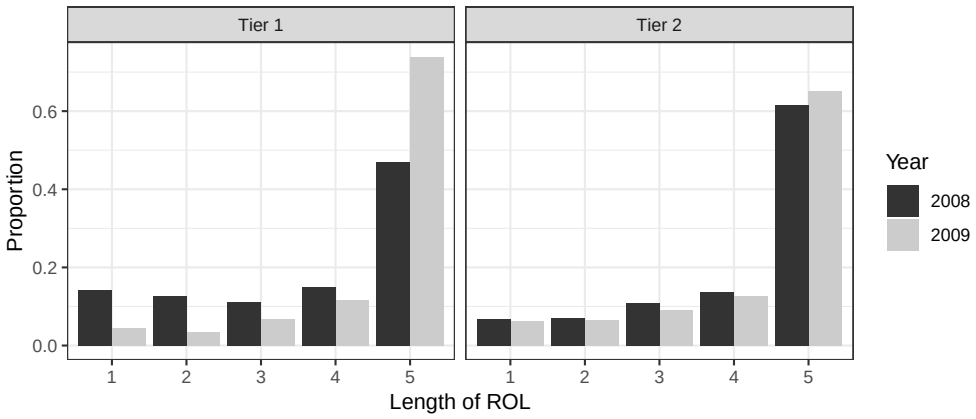


Figure 2: Distribution of the length of ROLs

Table 3: Effects of matching mechanisms on the length of rank-order lists (OLS)

Dependent variable:	Length of ROL		
	(1)	(2)	(3)
Y2009	0.073*	0.073*	0.076
	(0.041)	(0.041)	(0.049)
Tier1	-0.489**	-0.490**	-0.452**
	(0.201)	(0.211)	(0.210)
Y2009 $\times$ Tier1	0.724***	0.724***	0.708***
	(0.146)	(0.152)	(0.151)
Percentile Ranking		-0.784***	-0.776***
		(0.170)	(0.168)
Science			-0.306***
			(0.033)
Rural			-0.053
			(0.033)
Female			0.133***
			(0.024)
Constant	4.166***	4.557***	4.767***
	(0.038)	(0.083)	(0.097)
Observations	7,876	7,876	7,876
R-squared	0.021	0.053	0.070
Y2009 + Y2009 $\times$ Tier1	0.797***	0.797***	0.784***
	(0.181)	(0.188)	(0.194)
Tier1 + Y2009 $\times$ Tier1	0.235***	0.234***	0.257***
	(0.086)	(0.087)	(0.089)

Notes: Standard errors in parentheses are clustered at the high school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

fifth choice as insurance. Finally, the results in column (3) in Table 3 show that higher ranked students as well as those in the sciences tend to submit shorter lists while women tend to submit longer lists.

Next, we investigate whether students list more prestigious colleges as their top choices under PA, as predicted by theory. To compute our *prestige* index, we rank colleges from most (1) to least (n) prestigious, as measured by the average scores of students within a tier or track market. We calculate these rankings separately for 2006 and 2007 and then average the two to obtain a final prestige score for each college. These ranks are then normalized to range from 0 (most prestigious) to 1 (least prestigious) within each of the eight markets. Since not all colleges that admitted students in 2008 and 2009 did so in 2006 and 2007, observations with these colleges as top choices (2.2%) are dropped from our analysis. Compared to alternative measures, our prestige index utilizes the same data and statistics published and distributed to students and their parents by the Sichuan Educational Examination Authority in the *Gaokao Guide* (Sichuan: UESTC Press, 2009). While our prestige index is highly correlated with the published national rankings of colleges,[8] using the average score of admitted students provides a more accurate aggregation of students' revealed preferences for colleges compared to national rankings, because students without complete preferences over colleges often use cutoff and average test scores to assess a given college's prestige.[9]. In addition, national rankings may not necessarily reflect the local preferences of students in Sichuan, compared to our province-specific calculations.

Figure 3 presents the average prestige index for students' first-choice colleges by year and tier, with 0 (1) indicating the most (least) prestigious college. From Figure 3, we see that, on average, students choose more prestigious colleges in both tiers in 2009, compared to the 2008 choices, with a more pronounced increase for Tier 1 students.

We next examine the effect of the change from IA to PA on the prestige level of students' first choices. In this analysis, the dependent variable is the prestige level (from 0 to 1) of the student-reported top choice colleges. The independent variables (omitted) again include Y2009 (Y2008), Tier1 (Tier2), Y2009 $\times$ Tier1, Science (Humanities), Rural (Urban), Female (Male), and Percentile Ranking. The results in Table 4 show a negative and significant coefficient for our main treatment effect, Y2009 $\times$ Tier1, indicating that students list more prestigious colleges as their Tier 1 first choices in 2009, with a magnitude of 2.5% ($p < 0.05$). This result is consistent with the theoretical prediction that students are more likely to pick prestigious colleges as their top choice under PA since they are also able to include a safer choice in their ROL as well (Hypothesis 2). The effect, however, becomes weaker when control variables are added (column (3)). We further find that students' Tier 2 first choices in 2009 are ranked 3.5% higher than their corresponding rankings in 2008 ($p < 0.001$). It is not clear what drives this effect. Finally, the results in column (3) in Table 4 show that

[8]The correlation coefficient between the national ranking and our prestige index is 0.68 ($p < 0.001$, $n = 476$) for the science track, and 0.67 ($p < 0.001$, $n = 379$) for the arts track. We use 2009 Chinese College Rankings published by the Chinese Alumni Network as to obtain our national rankings. This data is chosen because it is the most complete published rankings encompassing more than 500 colleges each year.

[9]This is also the reason why top universities in China announce when they have a cutoff score higher than those of their rivals. See, e.g., https://cn.nytimes.com/education/20150701/c01sino-rivalry/en-us/. "If these students are taken by the competitors, then you'll be forced to lower your own cutoff score; once your cutoff score is lower than those of your competitors, you lose half of the battle of recruitment."

Table 4: Effects of matching mechanisms on the prestige of reported top choices (OLS)

Dependent variable:	Prestige index of top choice college		
	(1)	(2)	(3)
Y2009	-0.035***	-0.037***	-0.038***
	(0.011)	(0.012)	(0.012)
Tier1	0.115***	0.112***	0.109***
	(0.014)	(0.005)	(0.006)
Y2009 × Tier1	-0.025**	-0.021*	-0.019
	(0.012)	(0.013)	(0.013)
Percentile Ranking		-0.722***	-0.721***
		(0.026)	(0.026)
Science			0.024***
			(0.005)
Rural			0.043***
			(0.011)
Female			-0.019***
			(0.007)
Constant	0.403***	0.767***	0.722***
	(0.029)	(0.020)	(0.029)
Observations	7,706	7,706	7,706
R-squared	0.021	0.449	0.455
Y2009 + Y2009 × Tier1	-0.060***	-0.058***	-0.057***
	(0.009)	(0.006)	(0.006)
Tier1 + Y2009 × Tier1	0.091***	0.091***	0.090***
	(0.017)	(0.013)	(0.012)

Notes: A prestige index of 0 (1) indicates the most (least) prestigious college; Standard errors in parentheses are clustered at the high school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

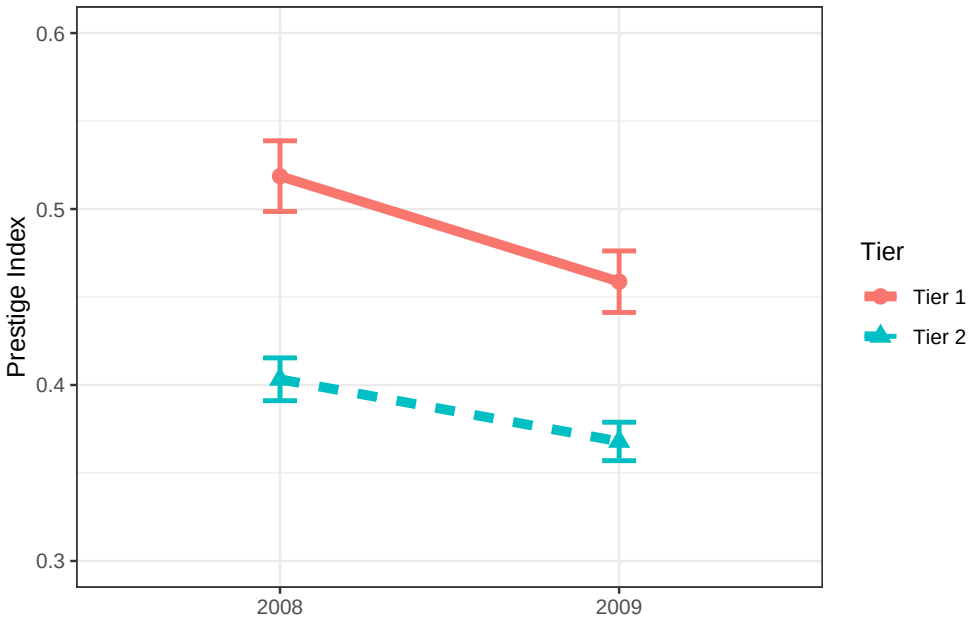


Figure 3: Average prestige of first-choice college by year and tier, with 0 (1) indicating the most (least) prestigious college; error bars indicate the 95% confidence interval of the mean.

higher ranked students as well as women list more prestigious colleges as their Tier 1 first choices in 2009 (-0.72 and -0.019, respectively, $p < 0.001$).

We now summarize our treatment effect of the type of mechanism on student preference ranking strategies:

Result 1. *Changing the Tier 1 admissions mechanism from IA to PA leads to an increase in the length of a student's rank-order list by approximately one more college, as well as a 2.5% increase in the prestige level of students' top-choice colleges.*

These empirical results are consistent with the theoretical prediction that students view the parallel mechanism (PA) as providing insurance or a fallback if they do not receive their ideal top choice, compared to the immediate acceptance mechanism (IA). Indeed, students appear to capitalize on the intuition that PA allows them to list more colleges and more prestigious colleges in the first tier without jeopardizing their admission chances to lower-ranked colleges within that tier.

5.2 Matching Outcomes

Next we investigate the effects of the type of mechanism and its subsequent behavioral changes on matching outcomes. First, we examine the effects of the type of mechanism change on the likelihood that a student is admitted by her reported first choice college. Recall that Hypothesis 3 predicts that IA will assign a larger

number of students to their reported first choices than PA, as students have an incentive to aim higher under PA.

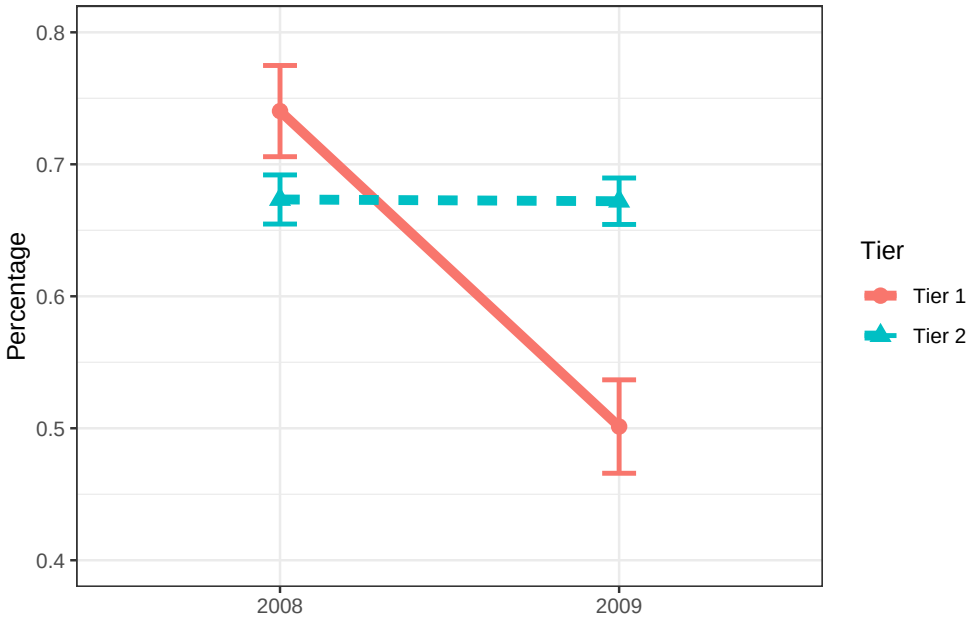


Figure 4: First choice accommodation rate by year and tier; error bars indicate the 95% confidence interval of the mean.

Figure 4 depicts the first choice accommodation rate by year and tier. As predicted by theory, we indeed see a drastic drop in the proportion of students admitted to their reported top-choice colleges in Tier 1 in 2009 (red solid line), in contrast to no change for Tier 2 admission rates (green dashed line). We next formally investigate this phenomenon through a regression analysis.

Table 5 reports the results of our regression analysis using three probit specifications: the effects of mechanism change on the likelihood of first choice accommodation (1), with student percentile ranking (2), and with demographic controls (3). The independent variables (omitted) again include Y2009 (Y2008), Tier1 (Tier2), Y2009 $\times$ Tier1, Science (Humanities), Rural (Urban), Female (Male), and Percentile Ranking.

Consistent with our theoretical prediction (Hypothesis 3), we find that the coefficient for our main treatment effect, Y2009 $\times$ Tier1, is negative and significant, indicating that students are 23.9 percentage points ($p < 0.01$) less likely to be admitted by their reported top choices in the Tier 1 admissions process in 2009, whereas the likelihood of being admitted by first-choice colleges in Tier 2 in 2009 does not change compared to the previous year (-0.001, $p > 0.10$). Additionally, looking at the covariates, we find that students from rural areas are 3.9 percentage points more likely to be admitted into their reported first choices under PA. Finally, we see that students with a one-percentile increase in their entrance exam scores increase their likelihood of being admitted by their reported first choice by 0.558% under PA ($p < 0.001$). Since PA incentivizes students to aim high, we also find a decrease in the acceptance rate of top choice colleges after

Table 5: Effects of matching mechanisms on first choice accommodation: Probit

Dependent variable:	Admitted to first choice		
	(1)	(2)	(3)
Y2009	-0.001	0.001	-0.001
	(0.023)	(0.021)	(0.022)
Tier1	0.067	0.077*	0.073*
	(0.052)	(0.039)	(0.039)
Y2009 $\times$ Tier1	-0.238***	-0.248***	-0.246***
	(0.050)	(0.047)	(0.048)
Percentile Ranking		0.558***	0.558***
		(0.017)	(0.016)
Science			0.038***
			(0.014)
Rural			0.039***
			(0.012)
Female			0.005
			(0.009)
Observations	6566	6566	6566
Y2009 + Y2009 $\times$ Tier1	-0.239***	-0.247***	-0.247***
	(0.042)	(0.041)	(0.040)
Tier1 + Y2009 $\times$ Tier1	-0.171***	-0.171***	-0.173***
	(0.026)	(0.027)	(0.026)

Notes: Coefficients report marginal effects; standard errors in parentheses are clustered at the high school level; standard errors for interaction terms are calculated using the Delta Method. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

the change to PA.

In addition to examining the effect of mechanism change on first-choice accommodation, we are interested in the performance of each mechanism in terms of matching stability. Recall that Hypothesis 4 predicts that PA will be more stable than IA. To measure stability, we first need to know students' preferences over colleges. In our study, we examine students' revealed preferences as indicated in their ROLs. This approach allows us to forego the assumption that students have identical preferences. With this measure, we assume that students preserve the preference order in their ROL, that is, they always list their more preferred colleges above their less preferred ones within the same choice band under PA, which is implied by Remark 3 in Chen and Kesten (2017).

We consider an outcome to be unstable in two situations. First, an outcome is considered unstable if a student in Tier 1 has a listed college above her admitted college (within the same tier) whose cutoff score is lower than her test score, indicating justified envy. Second, an outcome is considered unstable if a student ends up in a Tier 2 college or lower even though her test score is high enough to obtain admission into one of her listed Tier 1 colleges. For Tier 2 observations, the first condition is the same, whereas the second condition changes to receiving admission to a college below Tier 2.

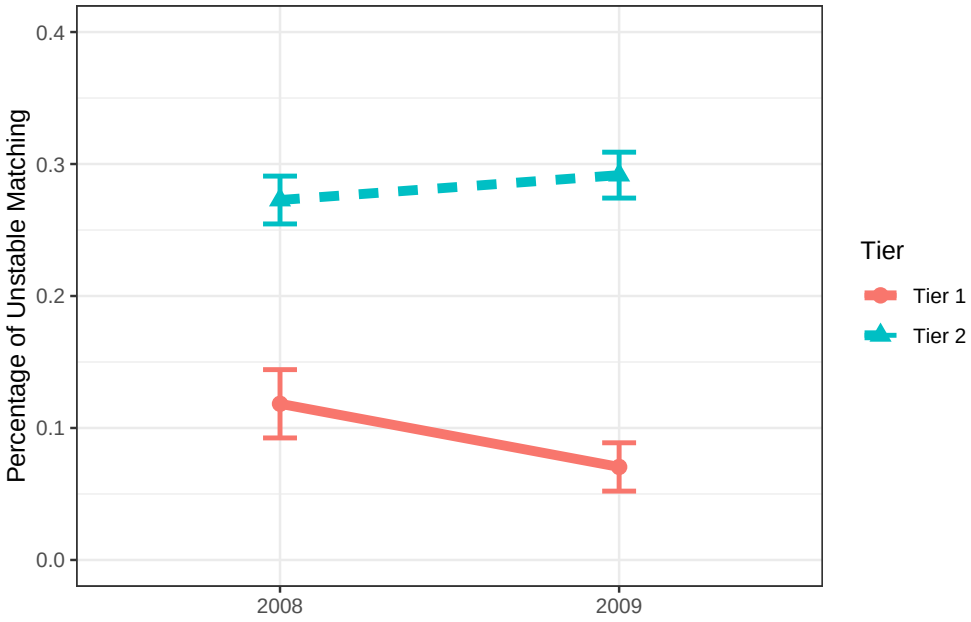


Figure 5: Proportion of unstable matching by year and tier; error bars indicate the 95% confidence interval of the mean.

Using this stability measure, we report the proportion of unstable matchings by year and tier in Figure 5. From 2008 to 2009, we see that the proportion of unstable matching decreases for Tier 1 students (solid red line), whereas that for Tier 2 students remains almost constant (green dashed line). To examine this effect further, we next conduct a regression analysis on the same outcome variable.

Table 6: Effects of matching mechanisms on stability: Probit.

Dependent variable:	Unstable Matching		
	(1)	(2)	(3)
Y2009	0.019	0.017	0.018
	(0.024)	(0.023)	(0.023)
Tier1	-0.154***	-0.157***	-0.156***
	(0.022)	(0.020)	(0.020)
Y2009 $\times$ Tier1	-0.067***	-0.063***	-0.064***
	(0.028)	(0.025)	(0.026)
Percentile Ranking		-0.326***	-0.327***
		(0.025)	(0.023)
Science			-0.016**
			(0.008)
Rural			-0.023**
			(0.009)
Female			-0.013
			(0.008)
Observations	6300	6300	6300
Y2009 + Y2009 $\times$ Tier1	-0.048***	-0.046***	-0.046***
	(0.014)	(0.010)	(0.010)
Tier1 + Y2009 $\times$ Tier1	-0.221***	-0.219***	-0.220***
	(0.017)	(0.020)	(0.020)

Notes: Coefficients report marginal effects; standard errors in parentheses are clustered at the high school level; standard errors for interaction terms are calculated using delta method. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 6 reports the results of our regression analysis of the effects of mechanism change on matching stability. The dependent variable here is whether the student's match is unstable. The independent variables (omitted) again include Y2009 (Y2008), Tier1 (Tier2), Y2009 $\times$ Tier1, Science (Humanities), Rural (Urban), Female (Male), and Percentile Ranking. From the results in Table 6, we see that the coefficient for our main treatment effect, Y2009 $\times$ Tier1, is negative and significant, indicating that the move to PA decreases the number of unstable outcomes by 6.7 percentage points ($p < 0.01$).

We summarize our matching outcome analysis findings below:

Result 2. *Changing the Tier 1 admissions mechanism from the IA to PA leads to a 24 percentage point decrease in the admissions students receive from their reported top-choice colleges, and a 6 percentage point decrease in the likelihood of unstable matchings.*

Our observed first choice accommodation result is consistent with theoretical predictions (Hypothesis 3): students are indeed more focused on getting into their reported first choices under IA. The stability result is also consistent with theoretical predictions (Hypothesis 4) in that PA results in fewer unstable outcomes. This latter result is robust to different measures of stability, including a cardinal measure of the distance between a student's exam score and the cutoff score (Appendix B).

To provide greater confidence in our findings, we next conduct a robustness test excluding the bottom 20% of Tier 1 students and the top 20% of Tier 2 students from our analysis. We do so to address the potential concern that the switch to PA in the Tier 1 process may impact the composition of students who participate in the Tier 2 process, as different mechanisms may leave different students unadmitted after the Tier 1 process concludes. This is important as our difference-in-differences estimates rely on the fact that the mechanism for Tier 2 does not change between 2008 and 2009. Excluding these students from our analyses yields similar results as those from our main analyses. Finally, we re-run our analyses using Tier 3 students as the control condition and find similar results. Appendices A and B summarize the results from these robustness checks.

6 Conclusion

The assignment of students to colleges is one of the most important education policy issues throughout the world, with significant social welfare and economic development implications attached to the process. In China alone, ten million high school students participate in the college admission process each year. Since 2001, the process of allocating available slots to students has changed from the immediate acceptance mechanism to various versions of the parallel mechanism. While the parallel mechanism has been shown to have numerous benefits on a theoretical level (Chen and Kesten, 2017), its benefits have been examined empirically mostly in a laboratory setting. By contrast, our study examines the effect of the parallel mechanism on student strategies and matching outcomes in a natural experiment using a unique data set with individual-level ranking strategies before and after the adoption of the new mechanism.

Specifically, we analyze a natural experiment using difference-in-differences estimators. Although some theoretical properties of matching mechanisms cannot be directly tested empirically due to the lack of students' true preferences, we can draw some analogies between the lab and the field using revealed preferences as seen in students' rank-ordered lists of their preferred colleges. We find that when the mechanism changes from IA to PA, students list better colleges as their first choices. We also find that students list more colleges in their rank-ordered lists under PA. These behavioral responses lead to more stable matching outcomes.

As college admissions reforms continue in China and other parts of the world (Westkamp, 2013; Braun et al., 2014), theoretical, experimental, and empirical analyses of ongoing reforms not only deepen our understanding of the science of market design, but also offer insights into how education and labor market policies should consider the choice of better mechanism in their implementation.

Supplemental Information

Appendix A Robustness checks

A.1 Excluding students at the boundary

Table A.1: Effects of matching mechanisms on the length of rank-ordered lists (OLS), excluding students in the bottom 20% test scores in Tier 1, and in the top 20% test scores in Tier 2.

Dependent variable:	Length of ROL		
	(1)	(2)	(3)
Y2009	0.104**	0.104**	0.107**
	(0.048)	(0.047)	(0.052)
Tier1	-0.506***	-0.574***	-0.552***
	(0.114)	(0.116)	(0.116)
Y2009 × Tier1	0.613***	0.612***	0.600***
	(0.144)	(0.144)	(0.144)
Percentile Ranking		0.342***	0.345***
		(0.094)	(0.092)
Science			-0.164***
			(0.029)
Rural			-0.060
			(0.040)
Female			0.077***
			(0.029)
Constant	4.433***	4.296***	4.432***
	(0.053)	(0.078)	(0.077)
Observations	6,295	6,295	6,295
R-squared	0.033	0.039	0.047
Y2009 + Y2009 × Tier1	0.717***	0.715***	0.707***
	(0.184)	(0.184)	(0.188)
Tier1 + Y2009 × Tier1	0.107***	0.038	0.048
	(0.037)	(0.041)	(0.041)

Notes: Standard errors in parentheses are clustered at the high school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Compared to the original result in Table 3, the average tier effect remains similar. The coefficient for the interaction becomes smaller, indicating that the difference in differences of the length of the ROLs becomes

smaller. The sign of the coefficient of percentile ranking changes. This may be caused by the difference in this variable's range after 20% of students are excluded based on their ranking of test scores (either bottom or top).

Table A.2: Effects of matching mechanisms on the prestige of reported first choices (OLS), excluding students in the bottom 20% test scores in Tier 1, and in the top 20% test scores in Tier 2.

Dependent variable:	Prestige of first-choice college		
	(1)	(2)	(3)
Y2009	-0.044***	-0.045**	-0.046***
	(0.016)	(0.017)	(0.017)
Tier1	0.006	0.158***	0.154***
	(0.008)	(0.005)	(0.006)
Y2009 × Tier1	-0.025*	-0.019	-0.015
	(0.015)	(0.018)	(0.018)
Percentile Ranking		-0.783***	-0.783***
		(0.017)	(0.017)
Science			0.025***
			(0.007)
Rural			0.048***
			(0.013)
Female			-0.021**
			(0.009)
Constant	0.483***	0.799***	0.750***
	(0.020)	(0.015)	(0.027)
Observations	6,154	6,154	6,154
R-squared	0.007	0.352	0.359
Y2009 + Y2009 × Tier1	-0.069***	-0.064***	-0.062***
	(0.008)	(0.006)	(0.007)
Tier1 + Y2009 × Tier1	-0.019	0.139***	0.139***
	(0.015)	(0.015)	(0.014)

Notes: Standard errors in parentheses are clustered at the high school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Compared to the original result in Table 4, the previous weakly significant coefficient of the interaction is no longer significant, which may similarly be explained by the change in variable range.

Table A.3: Effects of matching mechanisms on first choice accommodation: Probit, excluding students in the bottom 20% test scores in Tier 1, and in the top 20% test scores in Tier 2.

Dependent variable:	Admitted to first choice		
	(1)	(2)	(3)
Y2009	-0.012	-0.009	-0.012
	(0.028)	(0.026)	(0.026)
Tier1	0.227***	0.156***	0.152***
	(0.027)	(0.024)	(0.023)
Y2009 × Tier1	-0.256***	-0.292***	-0.288***
	(0.039)	(0.041)	(0.041)
Percentile Ranking		0.516***	0.517***
		(0.022)	(0.021)
Science			0.037**
			(0.016)
Rural			0.047***
			(0.013)
Female			-0.001
			(0.012)
Observations	5260	5260	5260
Y2009 + Y2009 × Tier1	-0.268***	-0.301***	-0.300***
	(0.021)	(0.029)	(0.027)
Tier1 + Y2009 × Tier1	-0.029	-0.136***	-0.136***
	(0.030)	(0.030)	(0.029)

Notes: Standard errors in parentheses are clustered at the high school level; reporting marginal effects; standard errors for interaction terms are calculated using delta method. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Compared to the original result in Table 5, the most obvious change is the coefficient for the tier effect. This is however not surprising given that we exclude those students least likely to be admitted to their first choice for Tier 1, and most likely to be admitted to their first choice for Tier 2.

Table A.4: Effects of matching mechanisms on stability: Probit; excluding 20% at the margin

Dependent variable:	Unstable Matching		
	(1)	(2)	(3)
Y2009	0.035	0.033	0.034
	(0.029)	(0.028)	(0.028)
Tier1	-0.198***	-0.161***	-0.160***
	(0.028)	(0.025)	(0.025)
Y2009 × Tier1	-0.071***	-0.074***	-0.075***
	(0.031)	(0.029)	(0.029)
Percentile Ranking		-0.301***	-0.302***
		(0.032)	(0.031)
Science			-0.012*
			(0.008)
Rural			-0.026**
			(0.012)
Female			-0.008
			(0.010)
Observations	5029	5029	5029
Y2009 + Y2009 × Tier1	-0.036**	-0.041***	-0.041***
	(0.014)	(0.015)	(0.015)
Tier1 + Y2009 × Tier1	-0.269***	-0.235***	-0.235***
	(0.024)	(0.025)	(0.024)

Notes: Standard errors in parentheses are clustered at the high school level; reporting marginal effects; standard errors for interaction terms are calculated using delta method. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Compared to the original result in Table 6, this result remains robust and similar.

A.2 Comparing Tier 1 and Tier 3

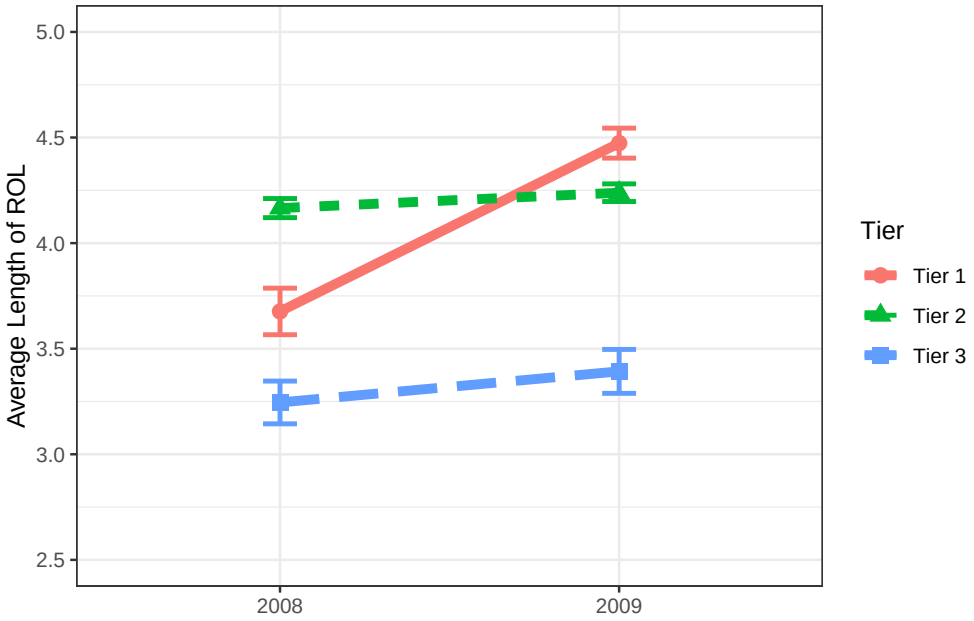


Figure A.1: Average length of ROLs by year and tier; error bars indicate the 95% confidence interval of the mean.

Table A.5: Effects of matching mechanisms on the length of rank-ordered lists (OLS), Tier 1 vs 3.

Dependent variable:	Length of ROL		
	(1)	(2)	(3)
Y2009	0.147	0.147	0.163
	(0.132)	(0.131)	(0.127)
Tier1	0.431***	0.431***	0.523***
	(0.139)	(0.135)	(0.144)
Y2009 × Tier1	0.650***	0.650***	0.628***
	(0.154)	(0.154)	(0.156)
Percentile Ranking		0.163	0.150
		(0.113)	(0.109)
Science			-0.284***
			(0.041)
Rural			-0.137***
			(0.043)
Female			-0.013
			(0.039)
Constant	3.245***	3.164***	3.425***
	(0.115)	(0.070)	(0.092)
Observations	3,229	3,229	3,229
R-squared	0.107	0.108	0.117
Y2009 + Y2009 × Tier1	0.797***	0.797***	0.792***
	(0.181)	(0.180)	(0.185)
Tier1 + Y2009 × Tier1	0.081***	0.081***	0.152***
	(0.061)	(0.058)	(0.065)

Notes: Standard errors in parentheses are clustered at the high school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Compared to the original result in Table 3, length of ROL becomes generally shorter when compared to Tier 2, but the difference-in-differences estimator remains robust.

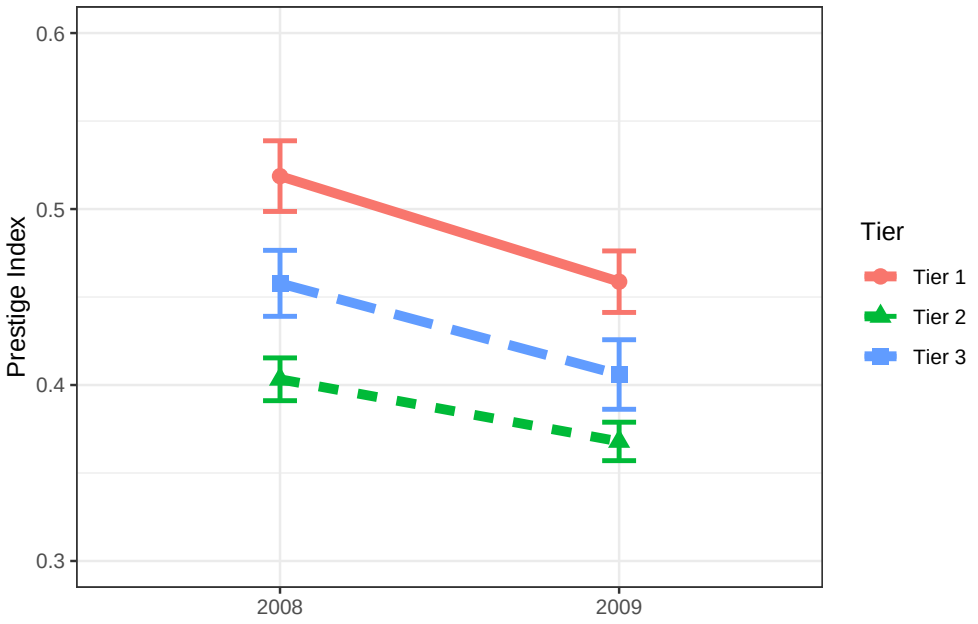


Figure A.2: Average first choice prestige by year and tier, with 0 (1) being the most (least) prestigious college; error bars indicate the 95% confidence interval of the mean.

Table A.6: Effects of matching mechanisms on the prestige of reported first choices (OLS), Tier 1 vs 3.

Dependent variable:	Prestige of first-choice college		
	(1)	(2)	(3)
Y2009	-0.052***	-0.054***	-0.053***
	(0.009)	(0.007)	(0.008)
Tier1	0.061**	0.061***	0.063***
	(0.026)	(0.016)	(0.016)
Y2009 × Tier1	-0.008	-0.006	-0.006
	(0.010)	(0.008)	(0.008)
Percentile Ranking		-0.402***	-0.398***
		(0.031)	(0.031)
Science			-0.013
			(0.011)
Rural			0.033***
			(0.011)
Female			0.002
			(0.010)
Constant	0.458***	0.658***	0.638***
	(0.013)	(0.011)	(0.009)
Observations	3,081	3,081	3,081
R-squared	0.02	0.202	0.205
Y2009 + Y2009 × Tier1	-0.060***	-0.059***	-0.059***
	(0.009)	(0.007)	(0.007)
Tier1 + Y2009 × Tier1	0.053**	0.055***	0.057***
	(0.022)	(0.010)	(0.011)

Notes: Standard errors in parentheses are clustered at the high school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Compared to the original result in Table 4, the previous weakly significant coefficient of the interaction is no longer significant, which may similarly be explained by the change in variable range.

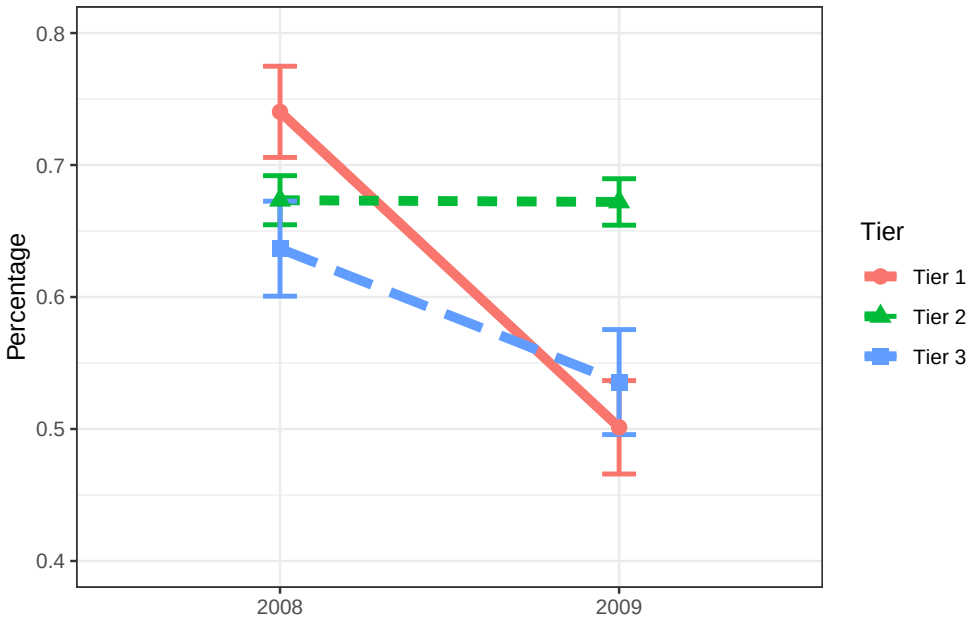


Figure A.3: First choice accommodation rate by year and tier; error bars indicate the 95% confidence interval of the mean.

Table A.7: Effects of matching mechanisms on first choice accommodation: Probit, Tier 1 vs 3.

Dependent variable:	Admitted to first choice		
	(1)	(2)	(3)
Y2009	-0.101***	-0.098***	-0.099***
	(0.025)	(0.020)	(0.020)
Tier1	0.104**	0.115***	0.098***
	(0.049)	(0.029)	(0.024)
Y2009 × Tier1	-0.138***	-0.149***	-0.148***
	(0.035)	(0.034)	(0.033)
Percentile Ranking		0.640***	0.639***
		(0.035)	(0.034)
Science			0.043**
			(0.019)
Rural			0.043***
			(0.011)
Female			-0.020*
			(0.011)
Observations	2681	2681	2681
Y2009 + Y2009 × Tier1	-0.239***	-0.247***	-0.247***
	(0.042)	(0.041)	(0.040)
Tier1 + Y2009 × Tier1	-0.034*	-0.034**	-0.050***
	(0.021)	(0.015)	(0.017)

Notes: Standard errors in parentheses are clustered at the high school level; reporting marginal effects; standard errors for interaction terms are calculated using delta method. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Compared to the original result in Table 5, this result remains robust, although generally fewer Tier 3 students get into their top choice than under Tier 2.

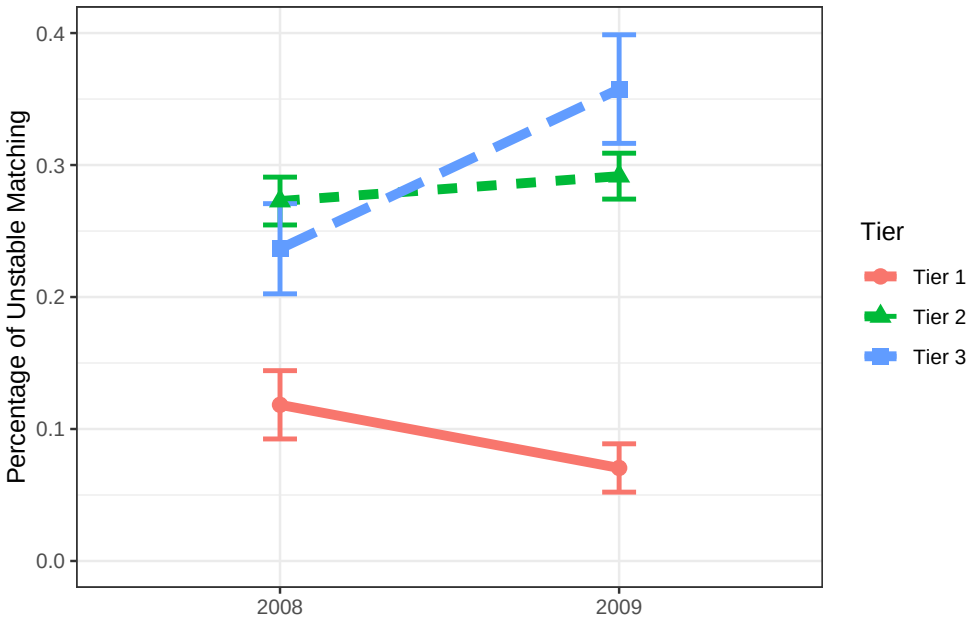


Figure A.4: Proportion of unstable matching by year and tier; error bars indicate the 95% confidence interval of the mean.

Table A.8: Effects of matching mechanisms on stability: Probit; Tier 1 vs 3

Dependent variable:	Unstable Matching		
	(1)	(2)	(3)
Y2009	0.121***	0.117***	0.118***
	(0.029)	(0.027)	(0.026)
Tier1	-0.118***	-0.127***	-0.118***
	(0.024)	(0.023)	(0.024)
Y2009 × Tier1	-0.169***	-0.163***	-0.165***
	(0.031)	(0.029)	(0.028)
Percentile Ranking		-0.162***	-0.162***
		(0.027)	(0.027)
Science			-0.018
			(0.015)
Rural			-0.018
			(0.015)
Female			0.010
			(0.009)
Observations	2471	2471	2471
Y2009 + Y2009 × Tier1	-0.048***	-0.046***	-0.047***
	(0.014)	(0.011)	(0.011)
Tier1 + Y2009 × Tier1	-0.287***	-0.290***	-0.283***
	(0.017)	(0.015)	(0.016)

Notes: Standard errors in parentheses are clustered at the high school level; reporting marginal effects; standard errors for interaction terms are calculated using delta method. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Compared to the original result in Table 6, the magnitude of the interaction coefficient becomes even larger.

Appendix B Alternative stability measures

In this section, we provide two alternative measures of stability.

B.1 Stability based on approximated preferences

In the context of centralized college admissions, we characterize a matching outcome to be more stable if the proportion of students a student exhibits justified envy towards is lower. In our context, student 1 exhibits *justified envy* toward student 2 if student 2 scores lower in the college entrance exams, but is admitted into

a college that is more preferred by student 1. Using this measure, we assume homogeneous preferences by all students, approximated by the ranking construct we used for the regressions presented in Table 4. Since matching is carried out separately by track, year and tier, we divide our sample into 8 markets, and calculate justified envy within the confinement of each individual market. For each student, we calculate the percentage of students, out of all students in each market, toward whom the student exhibits justified envy, and use this percentage as the dependent variable. A higher percentage means the matching is more unstable. The results are reported in Table B.9, with respective robustness checks in Table B.10 and B.11.

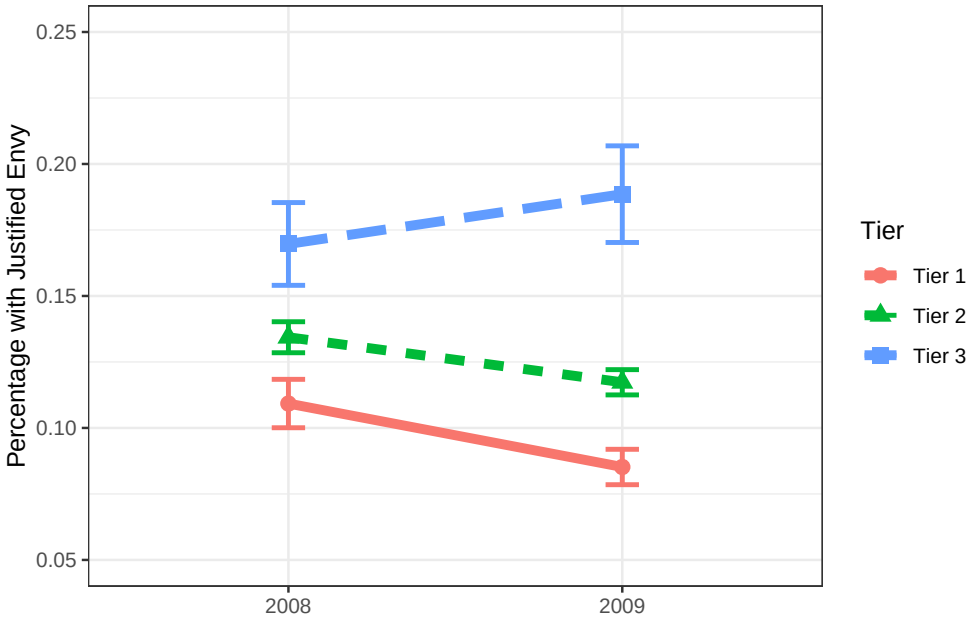


Figure B.5: Average percentage of students with justified envy by year and tier; error bars indicate the 95% confidence interval of the mean.

Table B.9: Effects of matching mechanisms on stability: OLS

Dependent variable:	Percentage Justified Envy		
	(1)	(2)	(3)
Y2009	-0.017***	-0.017***	-0.018***
	(0.002)	(0.002)	(0.002)
Tier1	-0.025***	-0.029***	-0.030***
	(0.006)	(0.006)	(0.006)
Y2009 $\times$ Tier1	-0.007***	-0.006***	-0.006***
	(0.002)	(0.002)	(0.002)
Percentile Ranking		0.087***	0.087***
		(0.009)	(0.009)
Science			0.007*
			(0.003)
Rural			0.006**
			(0.003)
Female			-0.010***
			(0.003)
Constant	0.134***	0.090***	0.084***
	(0.005)	(0.006)	(0.006)
Observations	6,061	6,061	6,061
R-squared	0.014	0.051	0.054
Y2009 + Y2009 $\times$ Tier1	-0.024***	-0.024***	-0.023***
	(0.003)	(0.004)	(0.003)
Tier1 + Y2009 $\times$ Tier1	-0.032***	-0.035***	-0.036***
	(0.007)	(0.008)	(0.008)

Notes: Standard errors in parentheses are clustered at the high school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B.10: Effects of matching mechanisms on stability: OLS, excluding students in the bottom 20% test scores in Tier 1, and in the top 20% test scores in Tier 2.

Dependent variable:	Percentage Justified Envy		
	(1)	(2)	(3)
Y2009	-0.017***	-0.017***	-0.018***
	(0.003)	(0.003)	(0.003)
Tier1	-0.015*	-0.049***	-0.051***
	(0.008)	(0.009)	(0.009)
Y2009 $\times$ Tier1	-0.010**	-0.010**	-0.009**
	(0.004)	(0.004)	(0.004)
Percentile Ranking		0.165***	0.166***
		(0.010)	(0.009)
Science			0.013***
			(0.004)
Rural			0.005*
			(0.003)
Female			-0.006**
			(0.002)
Constant	0.136***	0.068***	0.057***
	(0.005)	(0.006)	(0.006)
Observations	4,883	4,883	4,883
R-squared	0.011	0.11	0.113
Y2009 + Y2009 $\times$ Tier1	-0.026***	-0.027***	-0.027***
	(0.004)	(0.005)	(0.005)
Tier1 + Y2009 $\times$ Tier1	-0.024**	-0.059***	-0.060***
	(0.011)	(0.011)	(0.011)

Notes: Standard errors in parentheses are clustered at the high school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B.11: Effects of matching mechanisms on stability: OLS, Tier 1 vs 3

Dependent variable:	Percentage Justified Envy		
	(1)	(2)	(3)
Y2009	0.019**	0.022**	0.022**
	(0.009)	(0.009)	(0.009)
Tier1	-0.060***	-0.064***	-0.067***
	(0.009)	(0.012)	(0.013)
Y2009 × Tier1	-0.043***	-0.046***	-0.045***
	(0.008)	(0.007)	(0.007)
Percentile Ranking		0.188***	0.188***
		(0.029)	(0.029)
Science			0.006
			(0.005)
Rural			0.001
			(0.006)
Female			-0.007
			(0.006)
Constant	0.170***	0.069***	0.069***
	(0.008)	(0.019)	(0.019)
Observations	2,178	2,178	2,178
R-squared	0.076	0.192	0.193
Y2009 + Y2009 × Tier1	-0.024***	-0.023***	-0.023***
	(0.003)	(0.004)	(0.004)
Tier1 + Y2009 × Tier1	-0.103***	-0.109***	-0.112***
	(0.006)	(0.010)	(0.011)

Notes: Standard errors in parentheses are clustered at the high school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The results in the above three tables show that our stability findings are robust to this alternative stability measure.

B.2 Stability based on the distance between exam scores and cutoff scores

The measure in this section uses the distance between each student's test score and the cutoff score of his admitted college (i.e., "wasted" score). This measure is based on the following assumption: all else being equal, the larger the difference between these two scores, the more likely it is that there exists one or more schools that (1) a student lists but does not get into or (2) he does not list under more risky mechanisms due to strategic concerns. Point (2) captures student preference unobserved in our revealed preferences measure.

Both test scores and cutoff scores are standardized to Z-scores using the means and standard deviations of all students of each year and academic track, to account for any difference in exam difficulty. The results are reported in Table B.12, with respective robustness checks in Table B.13 and B.14.

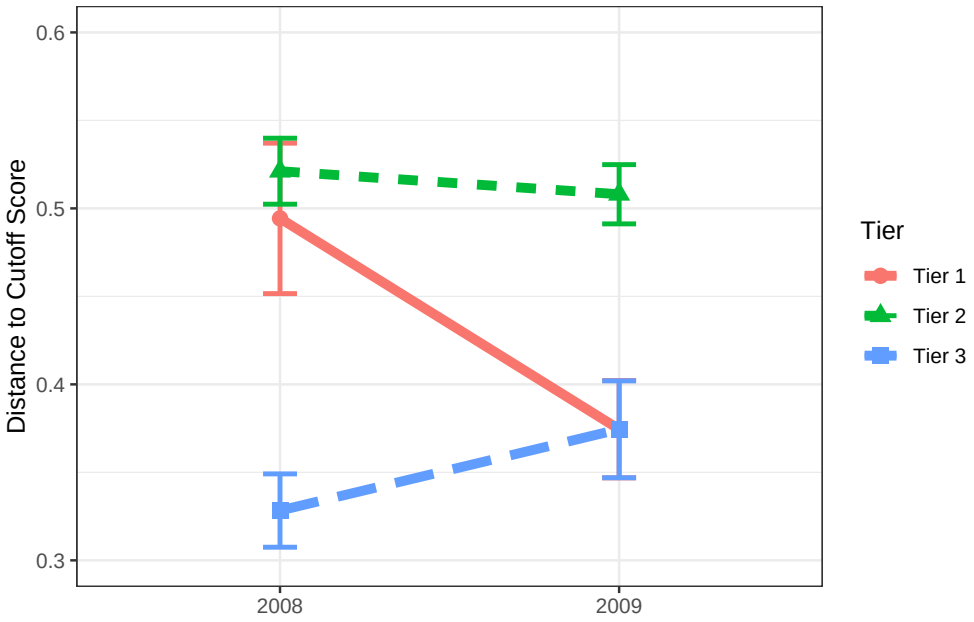


Figure B.6: Average distance between test score and cutoff score by year and tier; error bars indicate the 95% confidence interval of the mean.

Table B.12: Effects of matching mechanisms on stability: OLS.

Dependent variable:	Distance		
	(1)	(2)	(3)
Y2009	-0.013	-0.027**	-0.028**
	(0.010)	(0.012)	(0.012)
Tier1	-0.027***	-0.013	-0.015
	(0.006)	(0.018)	(0.017)
Y2009 × Tier1	-0.107***	-0.091***	-0.090***
	(0.021)	(0.028)	(0.028)
Percentile Ranking		0.694***	0.695***
		(0.011)	(0.011)
Science			0.029***
			(0.009)
Rural			0.001
			(0.013)
Female			0.013**
			(0.007)
Constant	0.521***	0.154***	0.126***
	(0.025)	(0.014)	(0.015)
Observations	5,566	5,566	5,566
R-squared	0.012	0.223	0.223
Y2009 + Y2009 × Tier1	-0.120***	-0.118***	-0.118***
	(0.017)	(0.021)	(0.021)
Tier1 + Y2009 × Tier1	-0.134***	-0.104***	-0.105***
	(0.019)	(0.014)	(0.014)

Notes: Standard errors in parentheses are clustered at the high school level; reporting marginal effects. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B.13: Effects of matching mechanisms on stability: OLS; excluding students in the bottom 20% test scores in Tier 1, and in the top 20% test scores in Tier 2.

Dependent variable:	Distance		
	(1)	(2)	(3)
Y2009	0.004	-0.009	-0.010
	(0.011)	(0.011)	(0.011)
Tier1	0.086***	-0.051***	-0.054***
	(0.009)	(0.015)	(0.014)
Y2009 × Tier1	-0.146***	-0.134***	-0.133***
	(0.015)	(0.017)	(0.017)
Percentile Ranking		0.784***	0.788***
		(0.020)	(0.020)
Science			0.045***
			(0.012)
Rural			0.005
			(0.014)
Female			0.020**
			(0.008)
Constant	0.427***	0.098***	0.052***
	(0.017)	(0.005)	(0.017)
Observations	4,328	4,328	4,328
R-squared	0.009	0.245	0.248
Y2009 + Y2009 × Tier1	-0.142***	-0.143***	-0.143***
	(0.014)	(0.015)	(0.015)
Tier1 + Y2009 × Tier1	-0.060***	-0.185***	-0.187***
	(0.013)	(0.008)	(0.008)

Notes: Standard errors in parentheses are clustered at the high school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B.14: Effects of matching mechanisms on stability: OLS; Tier 1 vs 3

Dependent variable:	Distance		
	(1)	(2)	(3)
Y2009	0.046***	0.061***	0.056***
	(0.014)	(0.010)	(0.012)
Tier1	0.166***	0.193***	0.175***
	(0.023)	(0.014)	(0.015)
Y2009 × Tier1	-0.166***	-0.179***	-0.174***
	(0.026)	(0.026)	(0.028)
Percentile Ranking		0.513***	0.512***
		(0.015)	(0.015)
Science			0.056***
			(0.019)
Rural			-0.016
			(0.017)
Female			-0.006
			(0.007)
Constant	0.328***	0.040***	0.026
	(0.012)	(0.010)	(0.021)
Observations	2,228	2,228	2,228
R-squared	0.025	0.154	0.158
Y2009 + Y2009 × Tier1	-0.120***	-0.119***	-0.118***
	(0.017)	(0.020)	(0.021)
Tier1 + Y2009 × Tier1	0.000	0.013	0.001
	(0.032)	(0.016)	(0.015)

Notes: Standard errors in parentheses are clustered at the high school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table B.12 to B.14 reports the effects of mechanism change on matching stability using our score distance measures. For each student, we calculate the z-score (the number of standard deviations away from the mean) between his test score and the cutoff score of his admitted college, and use this measure as the dependent variable. A larger z-score means the matching is more unstable. The independent variables (omitted) again include Y2009 (Y2008), Tier1 (Tier2), Y2009 × Tier1, Science (Humanities), Rural (Urban), Female (Male), and Percentile Ranking. We find that the main treatment effect, Y2009 × Tier1, is negative and significant (-0.107, $p < 0.01$). This again shows that our stability results are robust to different measures.

References

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

Abdulkadiroğlu, A., Agarwal, N., and Pathak, P. A. (2017). The welfare effects of coordinated assignment: Evidence from the New York City high school match. *American Economic Review*, 107(12):3635–89.

Abdulkadiroğlu, A., Pathak, P. A., and Roth, A. E. (2005a). The New York City high school match. *American Economic Review*, 95(2):364–367.

Abdulkadiroğlu, A., Pathak, P. A., Roth, A. E., and Sönmez, T. (2005b). The Boston public school match. *American Economic Review*, 95(2):368–371.

Agarwal, N. and Somaini, P. (2018). Demand analysis using strategic reports: An application to a school choice mechanism. *Econometrica*, 86(2):391–444.

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

Braun, S., Dwenger, N., and Kübler, D. (2010). Telling the truth may not pay off: an empirical study of centralized university admissions in germany. *The BE Journal of Economic Analysis & Policy*, 10(1).

Braun, S., Dwenger, N., Kübler, D., and Westkamp, A. (2014). Implementing quotas in university admissions: An experimental analysis. *Games and Economic Behavior*, 85(0):232 – 251.

Calsamiglia, C., Fu, C., and Güell, M. (2019). Structural estimation of a model of school choices: The boston mechanism vs. its alternatives. *Journal of Political Economy*.

Calsamiglia, C., Haeringer, G., and Klijn, F. (2010). Constrained school choice: An experimental study. *American Economic Review*, 100(4):1860–74.

Carvalho, J. R. and Magnac, T. (2010). College choice and entry exams. Working paper.

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

Chen, Y. and Kesten, O. (2019). Chinese college admissions and school choice reforms: An experimental study. *Games and Economic Behavior*, 115:83 – 100.

Chen, Y. and Sönmez, T. (2006). School choice: An experimental study. *Journal of Economic Theory*, 127:202–231.

De Haan, M., Gautier, P. A., Oosterbeek, H., and van der Klaauw, B. (2015). The performance of school assignment mechanisms in practice. CEPR Discussion Paper No. DP10656.

Ergin, H. I. and Sönmez, T. (2006). Games of school choice under the boston mechanism. *Journal of Public Economics*, 90(1-2):215 – 237.

Fack, G., Grenet, J., and He, Y. (2019). Beyond truth-telling: Preference estimation with centralized school choice and college admissions. *American Economic Review*, 109(4):1486–1529.

Featherstone, C. and Niederle, M. (2013). Improving on strategy-proof school choice mechanisms: An experimental investigation. Technical report, Working Paper, Stanford University.

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

Hakimov, R. and Kübler, D. (2019). Experiments on matching markets: A survey. WZB Discussion Paper.

He, Y. (2014). Gaming the Boston school choice mechanism in Beijing. Rice University Working Paper.

Jiang, M. (2016). When do stable matching mechanisms fail? the role of standardized tests in college admissions. Technical report, Working paper.

Kagel, J. H. and 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.

Kapor, A., Neilson, C. A., and Zimmerman, S. D. (2018). Heterogeneous beliefs and school choice mechanisms. Technical report, National Bureau of Economic Research.

Kesten, O. (2006). On two competing mechanisms for priority-based allocation problems. *Journal of Economic Theory*, 127(1):155 – 171.

Ledyard, J. O. and Szakaly-Moore, K. (1994). Designing organizations for trading pollution rights. *Journal of Economic Behavior & Organization*, 25(2):167 – 196.

Lien, J. W., Zheng, J., and Zhong, X. (2016). Preference submission timing in school choice matching: testing fairness and efficiency in the laboratory. *Experimental Economics*, 19(1):116–150.

Milgrom, P. (2019). Auction market design: Recent innovations. *Annual Review of Economics*, 11(1):383–405.

Milgrom, P. R. (2004). *Putting Auction Theory to Work*. Cambridge University Press, New York.

Mongell, S. and Roth, A. E. (1991). Sorority rush as a two-sided matching mechanism. *American Economic Review*, 81:441–464.

Nie, H. (2007). A game theoretic analysis of college admissions. *China Economic Quarterly*, 6(3):899–916.

Niederle, M. and 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.

Pais, J. and Pintér, Á. (2008). School choice and information: An experimental study on matching mechanisms. *Games and Economic Behavior*, 64(1):303 – 328.

Pan, S. (2019). The instability of matching with overconfident agents. *Games and Economic Behavior*, 113:396 – 415.

Pathak, P. A. and Sönmez, T. (2008). Leveling the playing field: sincere and sophisticated players in the boston mechanism. *American Economic Review*, 98(4):1636–1652.

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

Rees-Jones, A. and Skowronek, S. (2018). An experimental investigation of preference misrepresentation in the residency match. *Proceedings of the National Academy of Sciences*, 115(45):11471–11476.

Roth, A. (1990). New physicians:a natural experiment in market organization. *Science*, 250:1524–1528.

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. (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. (2015). Experiments in market design. In Kagel, J. H. and Roth, A. E., editors, *The Handbook of Experimental Economics*, volume 2, pages 290–346. Princeton University Press, Princeton, New Jersey.

Roth, A. E., Sönmez, T., and Ünver, M. U. (2004). Kidney Exchange. *The Quarterly Journal of Economics*, 119(2):457–488.

Song, Y., Tomoeda, K., and Xia, X. (2019). Sophistication and cautiousness in college applications. Manuscript.

Wei, L. (2009). A design for college and graduate school admission. *China Economic Quarterly*, 9(1):349–362.

Westkamp, A. (2013). An analysis of the german university admissions system. *Economic Theory*, 53(3):561–589.

Wu, B. and Zhong, X. (2014). Matching mechanisms and matching quality: Evidence from a top university in china. *Games and Economic Behavior*, 84(0):196 – 215.