Matching Mechanisms, Justified Envy, and College Admissions Outcomes

Le Kang
Wei Ha
Yang Song*
Sen Zhou

December 2020

Abstract

Matching mechanisms are crucial in centralized college admissions. In China, provinces centrally assign millions of students to universities every year. In the past two decades, many provinces changed from the Immediate Acceptance (IA) mechanism to the parallel mechanism, a hybrid of IA and the Deferred Acceptance mechanism. Using administrative data on millions of students, we find that changing to parallel mechanism reduced undesirable matching outcomes, including justified envy measures, null admission, and retaking. We also find a nonlinear and surprisingly nonmonotonic relationship between the improvement in justified envy measures and parallel choice bandwidth. Finally, congestion in mid-ranked universities could explain the higher levels of justified envy and larger improvements from mechanism change for students ranked in the middle.

Keywords: College admission; Immediate Acceptance mechanism; Boston mechanism; parallel mechanism; stability; justified envy; retake.
JEL Classification Numbers: C78, D82, I23, I28.

*Le Kang and Sen Zhou, China Institute for Educational Finance Research, Peking University. Wei Ha, Graduate School of Education and Institute of Economics of Education, Peking University. Corresponding author: Yang Song, Department of Economics, Colgate University. 13 Oak Dr, Hamilton, NY 13346. ysong@colgate.edu.

1 Introduction

College admissions have important, long-lasting implications for students (Hoekstra, 2009; Li et al., 2012; Dillon and Smith, 2020). A growing body of literature has studied the importance of market design in the school choice and college admissions context (Abdulkadiroğlu and Sönmez, 2003; Ergin and Sönmez, 2006; Pathak and Sönmez, 2008). The Immediate Acceptance mechanism (IA), also known as the Boston Mechanism, is one of the most popular matching mechanisms adopted by school districts to assign K-12 seats to households and for centrally organized college admissions to assign college seats to students. Following studies criticizing IA on its manipulability and instability, many policymakers have replaced IA with the Deferred Acceptance mechanism (DA), one that is stable and strategy-proof (Gale and Shapley, 1962; Abdulkadiroğlu et al., 2005, 2009). Although many prior studies have explored the theoretical properties and provided experimental evidence that compares these matching mechanisms,[1] the empirical evidence remains sparse.[2]

This paper is the first to use empirical data to compare justified envy measures and student outcomes under two matching mechanisms.[3] We study the Chinese college admissions, one of the largest matching market in the world, in which millions of students compete for limited seats at hundreds of universities every year. The college entrance exams and matching assignments are centralized at the provincial level, and many provinces changed their matching mechanism from IA to the parallel mechanism, a hybrid of IA and DA mechanism, in the past two decades. The temporal and geographical variations in the matching mechanism provide a unique opportunity for estimating causal effects of matching mechanisms. In addition, our individual-level administrative data allow for distributional analyses.

Unlike experimental settings where the true preferences can be pre-specified, the true preferences of students in real life are rarely observed by researchers. Fortunately, the Chinese higher education system provides a context where students within each matching game at the track-province-year level have overall homogeneous preferences over Tier I universities. Further, we allow students to have heterogeneous preferences horizontally within a rank group (for example, Peking and Tsinghua in China, or Harvard and Yale in the U.S.) and as-

[1]See Chen and Sönmez (2006); Calsamiglia et al. (2010); Chen et al. (2016); Chen and Kesten (2017); Chen et al. (2018); Basteck and Mantovani (2018) for a few examples.

[2]To our knowledge, only a few studies empirically documented theoretical predictions using empirical data in the US context (Abdulkadiroğlu et al., 2017a,b) and in the Chinese context (Ha et al., 2020; Chen et al., 2020). Agarwal and Somaini (2018); Calsamglia et al. (2020); Kapor et al. (2018) have explored the welfare implications of matching mechanisms using data from Cambridge, Barcelona, and New Haven. A few other studies (Bo et al., 2019; Cao, 2020) estimate the effects of mechanisms changes on student-college mismatch.

[3]Student A justifiably envies student B if A has a higher priority order than B (such as a higher test score) but is assigned to a lower ranked school. More rigorous definition will be given in subsection 2.1.

sume homogeneous preferences over groups of universities.[4] The rank groups are categorized based on the admission cutoff scores, i.e. the equilibrium "price" of seats at universities. Based on this assumption that students share homogeneous preference over university rank groups, we use millions of individual-level data on Chinese college entrance exams and admission outcomes between 2005 and 2011 to construct eight measures for justified envy (JE): whether a student justifiably envies any student, how many students one envies, the maximum and sum of degree of JE on the college quality dimension, the maximum and sum of degree of JE on the student rankings dimension, and the maximum and sum of degree of JE combining both previous two dimensions. Results from a generalized difference-in-differences framework suggest that the parallel mechanism reduces all these justified envy measures and brings the matching outcome closer to a stable matching.

We further investigate how the magnitudes of effects change over time and the relationship between improvement in justified envy and the number of parallel choices (hereafter bandwidth) in the first choice band.[5] Results suggest that the effect sizes of mechanism change increase over time, especially for measures on degrees of envy, but taper off at a certain level. The manitude of effects does not increase linearly, and surprisingly not even monotonically, with the parallel choice bandwidth. The effect sizes for six parallel choices are smaller than those of five parallel choices, especially for measures related to number of students envied and the sums of degree of envy. This finding holds true when we use outcome measures normalized to the matching game level to adjust for the nonlinearity in the increase of justified envy with matching game sizes.

Finally, we find an inverted-U shape relationship between student priority order (CEE scores) and their justified envy levels. The highest-ranked and lowest-ranked students have lower levels of justified envy under both IA and parallel mechanisms than mid-ranked students. Although a low level of justified envy is perhaps not surprising for the lower ranked students given that one needs to have a higher test score to justifiably envy others, the comparison between highest-ranked and mid-ranked students is interesting. We find evidence supporting that congestion in mid-ranked universities may explain higher levels of justified envy and larger improvements from mechanism change for students ranked in the middle.

We study two additional student outcomes under the IA and parallel mechanisms using data between 2002 and 2010 from Ningxia, an autonomous region with a population of

[4]The homogeneous preference assumption implies that there is no way to improve Pareto efficiency, limiting our ability to compare efficiency across mechanisms.

[5]We focus on the first choice band because most schools in Tier I are filled after the first choiceband are considered and admission assignments are final. Therefore, in this context, a parallel mechanism with six parallel choices is roughly equivalent to the DA mechanism with six choices, since after the first choice band concludes, most universities in Tier I are filled.

six million people. Unlike the national data, which contain only those admitted to some university, the Ningxia data include all CEE takers, allowing us to observe students who were not admitted by any school. It also allows us to identify students who retook in the CEE in the following year. We find that the changing from IA to parallel mechanism reduced the proportion of students who were not assigned. It also reduced the proportion of students who were so unsatisfied with the matching outcome that they repeated a year of high school to take the CEEs again.

The paper contributes to several strands of literature. First, it adds to the literature on school choice and college admission. Many empirical studies focus on the effect of winning a school choice lottery on student outcomes Cullen et al. (2006); Abdulkadiroğlu et al. (2011); Dobbie and Fryer Jr (2011); Deming et al. (2014). Research on the matching mechanisms is mostly theoretical and experimental. Although experimental studies are helpful in understanding theoretical predictions and providing valuable insights, empirical evidence is much needed for external validity support (Levitt and List, 2007). One reason is that experimental studies include far fewer participants and fewer schools per match than real world situations. Prior experimental studies typically have 3 to 36 participants per matching, while centralized college admissions, such as China and Turkey, can have hundreds of thousands of participants per matching. Second, the stake facing students in an experimental environment is much lower than in the real world. This study compares empirical estimates with corresponding effect size found in prior experimental studies (Chen and Kesten, 2019) and provide evidence on the external validity of experimental findings in the school choice setting.

Empirical studies that compare stability and justified envy across mechanisms are scarce, as doing so requires high-quality administrative data and a suitable empirical setting that allows for causal identification. Abdulkadiroğlu et al. (2017b) use data from New Orleans and Boston to show that the Top Trading Cycle has significantly less justified envy than serial dictatorship. Ha et al. (2020) use university-level college admissions data and show that changing from IA to parallel mechanism increases stratification, which is measured by rank ranges and match index.[6] Although these two measures correlate closely with matching stability, they are coarse and utilize only the maximum and minimum student ranks from each university. Chen et al. (2020) use data from a Chinese county to study how student strategies and matching outcomes are affected by changing from IA to parallel mechanism. They find that students apply to more colleges and list more prestigious colleges first; they also find that the student?college matchings become more stable.

The rich data from Ningxia also allow us to study null-admission and retaking behavior,

[6]Match index is calculated by (rank range + 1)/quota and should equal to 1 in the unique stable matching under the assumption that both students and schools share homogeneous strict preferences.

adding additional outcomes to the mechanism comparison. Several recent studies structurally estimate the student preferences and the distribution of household strategic types to produce welfare comparisons across matching organizations and mechanisms (Abdulkadiroğlu et al., 2017a; Agarwal and Somaini, 2018; Calsamglia et al., 2020; Kapor et al., 2018). Agarwal and Somaini (2018) use data from several public school districts in Cambridge and nonparametrically identify household preferences. They conclude that IA is preferred to DA by the average household, by an equivalent of 0.08 miles shorter distance to school, assuming that households behave optimally. Calsamglia et al. (2020) use data from Barcelona public school system and show that a change from IA to DA would decrease the average welfare because there are more losers than winners from the change. These two studies are both in the context of double-sided heterogeneous preferences with K-12 school choice. In college admissions, priorities are more finely ordered based on a continuum of scores, and preferences are more homogeneous on both sides.[7] Although null admission and retaking are far from a full picture of student welfare, we provide a first piece of empirical evidence in the college admission context in this direction. We find contrasting results to Agarwal and Somaini (2018) and Calsamglia et al. (2020) that changing from IA to parallel generated improvement in student outcomes: a smaller proportion of students with clearly unsatisfying outcomes, i.e. non-admission and taking another year to retake. Although for many retakers, spending an additional year often improves their individual admission outcomes, this effort generates little human capital gains and is socially costly (Krishna et al., 2018), similar to the rat race in Akerlof (1976).

Lastly, this paper contributes to a growing body of literature on Chinese college admissions, assigning millions of students to colleges every year (Nie, 2007; Zhu, 2014; Chen and Kesten, 2017). In particular, Chen and Kesten (2017) characterize the parallel mechanism in a system of matching mechanisms in between IA and DA and show that the parallel mechanism is more stable than the IA mechanism. Several studies compare the ex-ante fairness comparison between pre-exam preference submission in combination with IA and post-exam preference submission in combination with the parallel mechanism (Wu and Zhong, 2014; Zhong et al., 2004; Li et al., 2016; Lien et al., 2016, 2017). A few empirical studies in the Chinese college admission context are either qualitative Shen et al. (2008); Hou et al. (2009), have outcome limitations Li et al. (2010); Ha et al. (2020), or are using data from one county or one university Chen et al. (2020); Wu and Zhong (2014). This study takes advantage of an administrative data on almost the universe of students admitted to universities in China across seven years and another administrative data with the universe of test-takers in one

[7] Distance to home and sibling attending the same school are the two main factors driving heterogeneous preferences in the K-12 school choice. Neither matters as much in the college admissions setting.

province across nine years. These unique data sets allow us to have detailed understanding of the impacts of matching mechanism on individuals.

The rest of the paper is organized as follows. Section 2 provides a brief theoretical background and policy background. Section 3 describes the data and our key outcome variables. Section 4 specifies our empirical strategy, Section 5 presents the results, and Section 6 discusses the robustness checks. Section 7 concludes and discusses the policy implications.

2 Background

2.1 Matching Mechanisms and Theoretical Predictions

The theoretical literature on matching focuses on three main properties: stability, strategy-proofness, and Pareto efficiency. The justified envy measures in our paper are tightly connected to stability, and the incidences of null school assignment and retaking are loosely related to Pareto efficiency under some assumptions. We define these two properties and summarize theoretical predictions here and will discuss how our measures connect to them in the data section.

A college admissions problem includes N students, denoted by $I = i_1, i_2, ...i_n, ...i_N$, each to be assigned to one school in $S = s_1, s_2, ...s_j, ...s_J \cup \emptyset$, where $J \geq 2$ and $\emptyset$ denotes a student's outside option, or the null school. Each university has a limited number of seats, or quota q_j. Denote the preference order for student i by P_i and the preference order for university s by $\succ_s$. A matching $m : I \rightarrow S$ is a list of assignments that assign each student to a school, including the null school, and allows no university to admit more than its quota. A matching is *stable* if it is non-wasteful and justifiable envy-free. In a *non-wasteful* matching, no student prefers a school with any remaining seat to their own assignment, or all schools are filled. Student i *justifiably envies* student j for school s if i would rather be assigned to school s, where some student j who has lower priority than i, is assigned. Student i and school s are a *blocking pair*. A matching is *Pareto efficient* if there is no other matching that makes all students at least as well off and at least one student strictly better off.

Three mechanisms are of central relevance to this paper, the Immediate Acceptance (or Boston) mechanism, the parallel mechanism, and the Deferred Acceptance (DA) mechanism. The IA mechanism operates as follows:

Step 1. For each school s, consider only those students who list it as their first choice. Up to q_s students with the highest s preference orders among them are assigned to school s.

Step k, $k \geq 2$. Consider the remaining students. For each school s with q_s^k available seats, consider only those students who have listed it as their kth choice. The q_s students with the highest preference orders by school s among them are assigned to school s.

The algorithm continues until all students exhaust their lists or get assigned to a school. The IA mechanism has been criticized by its lack of strategy-proofness and stability, which are achieved by the DA Mechanism (Gale and Shapley, 1962). The student-proposing DA works as follows: *Step 1.* For each school s, up to q_s applicants who listed school s as their top choice and have the highest preference orders by school s are *tentatively* assigned to school s. The remaining applicants are rejected.

Step k, $k \geq 2$. Each student rejected from a school at step $k-1$ is considered by the next school listed. For each school s, up to q_s students who have the highest preference orders for school s among *both* the new applicants *and* those tentatively on hold from the earlier step are tentatively assigned to school s. The remaining applicants are rejected.

The algorithm continues until no student proposal is rejected. The parallel mechanism are in-between IA and DA mechanism. Students list several "parallel" universities within each choice band. Within each choice band, it operates like DA; across choice bands, it operates like IA. The parallel mechanism lies somewhere between IA and DA: Under IA, every choice is final; under DA, every choice is temporary until all seats are filled. Here we describe a parallel mechanism with two parallel choices in each choice band:

Step 1. For each school s, up to q_s applicants who listed school s as their top choice and have the highest preference orders by school s are tentatively assigned to school s. The remaining applicants are rejected.

Step 2. Each student rejected from a school at step 1 is considered at their second choice school. For each school s, up to q_s students who have the highest preference orders by school s among *both* the new applicants *and* those tentatively on hold from step 1 are assigned to school s. The remaining applicants are rejected. All assignments are final after all students' two parallel choices in the first choice band are considered.

Step 3. Step 1 and Step 2 are repeated using the two parallel choices in the second choice band for students who have not yet been assigned to any school. Seats assigned during the previous choice band consideration are no longer available. The algorithm continues until no student proposal is rejected.

As discussed by Chen and Kesten (2017) in more details, these three types of mechanisms can be thought of as a family of application-rejection mechanisms, $\phi^e, e \in \{1, 2, ..., \infty\}$. When $e = 1$, ϕ^e is equivalent to the IA mechanism. When $2 \leq e < \infty$, ϕ^e is equivalent to the parallel mechanism. When $e = \infty$, ϕ^e is equivalent to the DA mechanism. Chen and Kesten (2017) show that as e, the parallel choice band increases by an integer of $k > 1$, the mechanism becomes less manipulable and more stable.[8] These results provide the theoretical foundation for our empirical analysis.

Abdulkadiroğlu et al. (2017b) show that justified envy can be minimized by the Top Trading Cycle (TTC) mechanism among the strategy-proof mechanisms.

2.2 Chinese College Admissions

The college admissions system in China is centralized at the provincial level. Students choose either STEM or Non-STEM track at the end of 10th grade, study different subjects in the last two years of high schools, and take the College Entrance Exams at the end of 12th grade. All students study and take the CEEs in Chinese, mathematics, and English. In addition to these three subjects, students in the STEM track study and get tested on physics, chemistry, and biology, while students in the Non-STEM track do so for politics, history, and geography.[9] Due to the subject differences, the CEEs and college admission process are separate for each track. Universities allocate quotas to each province for each track, and the provincial education bureau assigns these seats to students based on their CEE scores and submitted college preferences.

The matching process differs in two dimensions across provinces: timing of preference submission and matching mechanism. In the first dimension, all provinces used to have students submit their preferences before they take the CEEs (ex-ante preference submission). Over time, some provinces changed to preference submission after taking the CEEs but before knowing the scores (ex-interim preference submission), and some changed to preference submission after knowing the scores (ex-post preference submission). The left panel in Figure 1 shows the variations in the preference submission timing over time. Between 2005 and 2011, ten provinces changed from either ex-ante or ex-interim to ex-post submission. By 2011, most provinces have adopted ex-post preference submission.

Before 2003, all provinces used the Immediate Acceptance (IA) mechanism. Under IA, almost all Tier I universities are filled after considering the first choice, so it is rare for students to be admitted into second or third choice in the same tier. If a student aims too high and fails to be admitted to the first choice university, this student could end up being assigned to a university in the next Tier, or not being assigned to any university

[8] Chen and Kesten (2017) also hypothesize a more general result where the manipulability and stability changes monotonically with the parallel choice band and provide simulation analysis to support it.

[9] Some provinces tried to experiment with the track system and allow students not to limit their choices of subjects within a track, such as allowing them to study and get tested for politics, chemistry, and biology. During our sample period between 2005 and 2011, the traditional two-track system was adopted across provinces. Jiangsu, Guangdong in 2005 and 2006, and Liaoning in 2005 adopted an alternative track system, therefore we exclude them from our sample.

at all. These students often repeat another year of high school to go through the college admission process again. Because of the risks discussed above, some students were overly conservative in their choices and significantly undershoot, which resulted in undermatch. In 2003, Hunan Province was the first to adopt the parallel mechanism. Following the positive feedback in the early reform provinces, the Ministry of Education issued a recommendation in 2008, which generated arguably exogenous adoption in the parallel mechanism adoption. The right panel of Figure 1 shows that the majority of the take-up happened in 2008 and 2009, following the state recommendation. By 2011, only seven provinces were still using the IA mechanism in their Tier I university assignment.

3 Data and Summary Statistics

We use two student-level administrative datasets to investigate the changes in justified envy and student welfare related to the matching mechanisms: a national dataset and a provincial dataset. The national dataset allows us to make causal inference and compare across parallel choice bandwidths, while the data from Ningxia province allow us to compare student outcomes, including null assignment and repeating.

3.1 National Data and Justified Envy Measures

The national administrative dataset includes every student who took the College Entrance Examinations (CEE) and was admitted by any university in China from 2005 to 2011. It has information on the year and province in which a student took the CEE, the CEE track (STEM or non-STEM), the CEE score, and the university they were admitted to.

We focus on Tier I universities and students with CEE scores above the admission threshold of Tier I universities for a few reasons. First, in China, only Tier I universities admit students from most provinces in large numbers, which is important for us to compare the justified envy across provinces. Second, the main purpose of the parallel reform is to reduce the cases where a high-achieving student ends up mismatched to a low-ranked university. Finally, to define justified envy, we need to assume that within each province, students have homogenous preferences towards groups of universities, and this assumption is most likely satisfied when students make choices among elite institutions.

With these considerations, we first restrict our sample to students with CEE scores above the province-year-track specific Tier I universities and got admitted to a Tier I university. We also exclude students who attend a university in Hong Kong and those who were admitted through art, music, and sports programs because CEE score is not the sole criteria for those admissions.[10] Our final national sample includes 3.73 million students; among them, 1.3 million students took CEE before the parallel reform.

3.1.1 Homogeneous Preferences

Many of our outcome variables are built upon the concept of justified envy, to define which we need information on university preferences over students and that on student preferences over universities. All universities rank students by their College Entrance Examination scores, other than two exceptions. One exception is that a few colleges admit students to special programs such as art, music, acting, and sports with another interview or exam process that is also considered as part of their selection criteria. The other exception is "zizhu zhaosheng" policy, or college-specific bonus points policy, where a small number of students can earn bonus points (usually 10 to 20 out of 750 points) for a specific college. These bonus points are not applied to other universities, which generates heterogeneous preferences by universities in the centralized matching process. However, this policy is normally restricted by the Ministry of Education to less than 5% of the admission quota and only started to become more widely adopted after 2011, the end of our sample period.[11]

Student preferences are relatively homogeneous due to the highly stratified higher education system and mostly single dimensional comparison among universities in China. Since the 1990s, the Chinese government poured resources into a selected group of universities, aiming to build world-class universities by concentrating resources on a smaller subset. University prestige and selectivity are highly stable over time, especially within each province, the level at which matching game takes place. In addition, unlike the American higher education system, where student have differing preferences over liberal arts colleges versus research universities, or universities with high rankings in popular sports, these factors essentially do not matter in the Chinese context. [12]

Other factors for heterogeneous preferences include distance to college, localized reputation, and localized employment premia differences due to alumni network. Fortunately, the

[10] Universities in Hong Kong often conduct an interview; art, music, and sports programs admit university-specific interviews and evaluations, which we do not observe their scores. The raw sample includes 11.74 million observations. Students with CEE scores above the Tier I admission cutoffs are 33.5% of the raw sample. Conditional on having a higher than Tier I cutoff score, about 9.12% were admitted to universities in Hong Kong and who were admitted through art, music, and sports programs.

[11] See `http://www.moe.gov.cn/srcsite/A15/moe_776/s3110/201111/t20111115_127339.html`, retrieved on 10/31/2018).

[12] All Tier I universities we focus on are public, large research universities. They are the only choices for high-achieving students, except for studying abroad. Those who plan to attend colleges abroad have to make their decision before the CCE takes place, so they do not participate in the CEE or the matching process either. Sports leagues receive very little attention in China. There is also no lower tuition rate for in-province students.

admission process is administered at the province level, which means that all students within each matching game are from the same province. Therefore, they are likely to have similar distances to universities, face similar employer premium for local universities if they would like to come home after college, and have been exposed to similar localized reputation. This strengthens our confidence in the homogeneous preference assumption.

Finally, some groups of universities have rankings and reputation so close to each other that students may be indifferent between them, or one student prefers one while a different student prefers another within the group (for example, Harvard and Yale in the American context, or Tsinghua and Peking in the Chinese context). Thus, instead of analyzing justified envy over school-specific ranks, we assign universities into rank groups based on students' revealed preference: average admission score percentile (AASP) within the track-province-year cell, i.e. the matching game, during the current year. This matching-specific construction allows university popularity to change over time within a province and allows university popularity to differ across provinces. To be consistent with the nature of higher education rankings in China, the group rank ranges is set to be wider as they get less selective. In particular, rank group one includes universities with AASP in the top 10%, and Rank Group two to four include universities with AASP between 10% to 30%, 30% to 60%, and 60% to 100%.

To test whether our results are robust to the rank group assignment to universities, we also use several other measures to define student homogeneous preferences over schools. Instead of assigning universities into four rank groups using the AASP in the current year, we use the AASP in the previous year and AASP through the entire sample period (2005-2011). Further, we use these three different ways to calculate AASPs to divide schools into eight rank groups, with the ranking percentile cutoffs at 5%, 10%, 20%, 30%, 45%, 60%, 80%. Finally, we also group universities based on the 985 Project and 211 Project university designation. Since the 1990s, the Chinese government poured resources into a selected group of universities, aiming to build world-class universities by concentrating resources on a smaller subset. The most prestigious group is the "985 Project" universities. There are 39 "985 Project" universities in total, with Peking University and Tsinghua University being the top two. The next level of selective universities is the "211 Project", including 112 universities. All the "985 Project" universities are "211 Project" universities, and all the "211 Project" universities are Tier I universities. Employers favor graduates from "985" and "211 Project" universities over other universities, the same way that graduates from prestigious flagship state universities and Ivy Leagues enjoy a signaling value from their degrees. We adopt this grouping of schools assuming student preferences across these groups to be: Peking and Tsinghua universities $>$"985 Project" universities $>$"211 Project" universities $>$Rest of Tier

I. We choose the local revealed preference based on AASP to be our main measure over this fixed Rank Group assignment, because there are often outliers in the fixed RG method that could cause misassignment of justified envy. For example, Jinan University in Guangdong Province is not a "985 Project" university, yet is highly regarded by employers and students in Guangdong and nearby provinces. Suppose that in Guangdong Province, student A is assigned to Jinan University with a CEE score of 640, and student B is assigned to a "985 Project" University with a CEE score of 630. If we use the "985-211" grouping method, we would deem that student A justifiably envies student B; however, in reality, it could be that student A genuinely prefers their assigned school, and no justified envy is present. We will test whether conclusions are similar across these different ways of assigning rank groups to universities.

3.1.2 Justified Envy Measures

For any student i, denote T_i the rank group of the student i's admitted university and r_i the ranking of student i. Consider student i as the focal student. i justifiably envies j if $r_i < r_j$ and $T_i > T_j$. Denote D^T_{ij} the degree of justified envy from i to any student j in the dimension of university rank group admitted. D^T_{ij} equals to zero if i does not justifiably envy j and equals to $T_i - T_j$ otherwise (note that $T_j < T_i$ is a necessary condition for i to envy j). Similarly, we can define the degree of justified envy in the dimension of student rankings, D^r_{ij}. It equals to zero if i does not justifiably envy j and equals to $r_j - r_i$ otherwise (note that $r_i < r_j$ is the other necessary condition for i to justifiably envy j). Finally, we construct a degree of envy combining both dimensions: $D^d_{ij} = \sqrt{(D^T_{ij})^2 + (D^r_{ij})^2}$, which is the diagonal distance between i and j when there is justified envy. To make these two dimensions comparable in data, where student rankings could differ in thousands while the university RG difference is at most three, we normalize student ranking to be corresponding to the university rank group. $r_i = 1$ when student i would be assigned to a university within Rank Group 1 in the perfectly stratified matching outcome, which assign students to colleges according to student and university ranking; $r_i = 2$ when student i would be assigned to a university within Rank Group 2 in this perfectly stratified matching, and so on.

For example, consider a matching outcome for eight students: A, B, C, D, E, F, G, H and four rank groups of universities, each with two seats. The priority rankings for student A through H are 1, 1, 3, 4, 5, 5, 7, 8, and their university rank groups admitted are 2, 1, 2, 4, 1, 3, 3, 4. Figure 2 visualizes this matching outcome, with student rank on the X-axis and university rank group admitted on the Y-axis. Since the necessary conditions for student i to justifiably envy student j include having a higher ranking than j and being assigned to a lower ranked university rank group than j, student i can only justifiably envy students who are to their lower right in the figure. In this example, we label all the justified envy pairs, with the start of the arrow being student i, the envier, and the end of the arrow being student j, the envied.

The matrix for degree of envy in terms of university rank group admitted $\mathbf{D^T}$ is

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The matrix for degree of envy in terms of student rank $\mathbf{D^r}$ is

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

We also calculate the combined degree of envy, which is the diagonal distance between i and j on Figure 2: $d_{ij}^d = \sqrt{(d_{ij}^T)^2 + (d_{ij}^r)^2}$. The matrix for degree of envy in terms of diagonal distance $\mathbf{D^d}$ is

$$\begin{pmatrix} 0 & 0 & 0 & 0 & \sqrt{17} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{5} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{10} & \sqrt{2} & \sqrt{10} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

For each matching game, we construct these three matrices, all of which are upper triangular because $r_i < r_j$ is a necessary condition for justified envy. We first extract two pieces of information for each row (i.e. each student i) from any of the three matrices. The first variable is whether student i justifiably envies anyone in the matching outcome, which is equivalent of whether row i has any non-zero element. The second variable is how many students student i justifiably envies, which is the number of non-zero elements in row i. Extracting these two variables from any of the three matrices yields the same values for these two measures.

We extract two additional variables measuring the maximum and sum for degree of envy along three dimensions from each of the three matrices. The maximum degree of envy of student i equals the largest value in row i, and the sum of degrees of envy for student i equals the sum of all elements in row i. Doing so for three matrices yielding six more justified envy measures along university rank group dimension, student rank dimension, and diagonal distance between any two students in a matching game. In total, we have eight measures. These measures for the illustrative example above are listed in Table 2.

So far, we adopt the standard definition for justified envy based on homogeneous preference assumptions. As we discussed in the previous subsection, the assumption that all students prefer one group of schools over another, as opposed to one school over another, is more likely to be true in reality. However, the kink created by the cutoffs between two rank groups could be problematic. For example, consider school A ranked 9.9 percentile and school B ranked 11 percentile. School A is assigned to RG 1, while school B is assigned to RG 2. However, some students may be indifferent between A and B, causing misassignment of justified envy. Therefore, we add a condition to justified envy in addition to $r_i < r_j$ and $T_i > T_j$: student i's university admitted must be 10% lower in ranking than that of j, $univpct_i - univpct_j > 10\%$.[13]

Although we use individual level data for our main analyses, we also collapse data to the matching game level to directly compare across games. To compare the matching outcomes at the game level across provinces, we need to normalize the measures because the degree of envy grows non-linearly with the size of a matching game. We do so by constructing the worst matching outcome, assigning the lowest ranked students to the highest university rank group up to its total quota, then the next lowest ranked students to the ssecond highest university rank group up to its total quota, and so on. This worst possible matching game gives us the largest justified envy measure for that matching game size. We then divide the sum of envy measures of the actual matching by the sum of envy measures of this worst possible matching to obtain the normalized envy measures at the matching game level.

[13]We will also test results without this additional condition, or a looser condition of $univpct_i - univpct_j > 5\%$.

3.2 Ningxia Data and Additional Student Outcomes

To complement the analyses using the national data, we use another individual-level administrative dataset between 2002 to 2010 from Ningxia to explore how the parallel reform affected student outcomes. This data set has two advantages. First, unlike the national data, which only include students who were admitted to a university, it includes all students who took the CEEs, regardless of whether a student was admitted by a university or not. Second, the Ningxia data contains richer demographic information at the student level, including birthday, gender, ethnic group, household registration status (urban or rural), and whether they were first-time CEE takers. After excluding 2,924 observations with zero CEE scores, the Ningxia sample includes 417,690 student-year observations, among whom 100,270 students took the CEE after the parallel reform.

Using Ningxia data, we construct two additional student outcomes. First, we create an indicator of whether a student is admitted to any university. Not being admitted to any university means that the choice universities listed by a student either prefer others over the student or were already filled up when it is their turn to consider the student. Some students not assigned to any university would attend a private vocational school, some start working, and others choose to repeat a year of high school and retake the CEE.

The second outcome measure directly looks at the retaking behavior. The decision to retake the entrance exam implies that the expected utility improvement in the admission outcome is greater than the cost of staying behind for an entire year studying the same materials under a high-pressure environment.[14] Consider a simple conceptual framework where student i's utility from being assigned to school s is denoted by u_i^s, and the cost of retaking is c_i^{retake}. A student would only repeat a year if $u_i^{s_{t+1}} - u_i^{s_t} > c_i^{retake}$. There is no obvious reason for student's expected improvement in admission outcome for the next year $u_i^{s_{t+1}}$ and the cost of retaking c_i^{retake} to change systematically and simultaneously with the mechanism reform. Thus, the retaking decision reveals information about student satisfaction with the admission outcome $u_i^{s_t}$. If anything, students may expect a larger improvement in the admission outcomes the following year, i.e. a better $u_i^{s_{t+1}}$, since the parallel mechanism provides a less risky environment than IA mechanism; if so, students may be *more* likely to repeat, biasing our results in the opposite direction. Although we are not able to compare the same set of students and their outcomes under two different mechanisms as in the definition of Pareto efficiency, we believe that students across cohorts are similar. A smaller percentage of students deciding to retake after the mechanism change suggests an improvement in student

[14] One could argue that repeating is not necessarily detrimental to a student if it improves their outcomes. That is beyond the scope of this paper. We try to focus on the assignment outcome and student evaluation of those outcomes.

satisfaction in matching outcomes.

3.3 Descriptive Statistics

The key explanatory variable of interest is the matching mechanism a particular province adopts in a given year. Table 1 shows the timing of reforms for each province. We also collect information about the timing of students submitting their university preferences. As the submission timing also could influence matching results, we control for the timing reforms in all regressions.

Table 3 presents the summary statistics of the national data, focusing on those who are in the Tier I universities and excluding those admitted to universities in Hong Kong or through art, music, or sports programs. There are 3.73 million observations in total between 2005 and 2011, 82% of which are in STEM track and 37% took the CEE in a province-year adopting the parallel mechanism. The majority of students (67%) submitted their preference lists after knowing their CEE scores.

The bottom panel of Table 3 presents summary statistics on justified envy measures. It may be surprising to see that 80.6% of students in our sample justifiably envies at least one other student, a very high percentage.[15] A high percentage of justified envy can be caused by just a handful of students with a very low CEE score being admitted into a high rank group (RG), because most students have a higher CEE score and justifiably envy those few students. See Figure 3 for an illustration, which shows the students admitted to four RGs in Shaanxi non-STEM track in 2009 and 2010. In 2009, Shaanxi adopted the IA mechanism, and 86% of students in the non-STEM track justifiably envy someone. The overlap in the score range across university rank groups was large. Justified envy happens whenever there is an overlap in the admitted students' rank range across two groups: the top student admitted to the lower rank group justifiably envies the last student admitted to the better rank group. Most students envy the student in the top left corner, who was admitted to the second rank group but with a low CEE score. Shaanxi adopted the parallel mechanism in 2010, and we can see from the right panel in Figure 3 that the overlap between rank groups got smaller. The lowest-priority student admitted to the second RG is significantly better, suggesting the parallel reform reduced justified envy and improved fairness and stability.

Moving onto the continuous measures of justified envy, a student justifiably envies 357 other students on average. For justified envy on the university rank group admitted, the mean of maximum degree of envy is 1.47, with the maximum being three since there are only

[15] By definition, students with lowest ranks and students admitted to the best rank group of universities never justifiably envy. If we drop the students with lowest ranks and students admitted to the best rank group, we find similar effects of changing matching mechanism on envy measures. These results are presented inTable A1.

four rank groups. The sum of justified envy on the university rank group admitted averages at 403. Turning to the other dimension of justified envy, the lowest ranked individual a student justifiably envies is ranked equivalent to 0.87 rank groups worth of ranking lower than themselves on average. When adding up the normalized student rank differences for all students envied, the average is around 168, lower than the sum degree of envy on the university rank group dimension.[16] Lastly, the diagonal degree of envy has, on average, a maximum degree of 1.74 and a sum degree of 456.

Table 4 presents the summary statistics at the individual level for the Ningxia data between 2002 and 2010. Among the 417,690 students, around 36.6% were not admitted to any universities, and 22.4% chose to retake CEE in the following year. The parallel reform happened in 2009 for Ningxia, and 26.4% of our sample individuals took the CEE after the reform. Around half of the students are male, 45% of them have urban hukou, and 77.7% of them are Han ethnicity.[17] Two thirds were in the STEM track.

4 Empirical Strategy

4.1 National Data Analyses

Exploiting the geographic and timing variations in the college admission mechanism reforms, we use a generalized difference-in-differences framework to identify the causal effect of changing from the IA to the parallel mechanism on justified envy measures.

$$\begin{aligned} Y_{ijst} &= \beta_0 + \beta_1 Parallel_{jt} + \beta_2 Ex-interim_{jt} + \beta_3 Ex-post_{jt} + \beta_4 rank_i \\ &\quad +\beta_5 N_{jst} + \alpha_j + \gamma_s + \delta_t + u_{ijst} \end{aligned} \tag{1}$$

Y_{ijst} indicates the set of justified envy measure for student i in province j track s year t. $Parallel_{jt}$ is the independent variable of interest, indicating whether province j adopted the parallel mechanism in year t. If the parallel mechanism reduces the incidences and degrees of justified envy, the coefficients of β_1 should be negative and statistically significant. $Ex-interim_{jt}$ and $Ex-post_{jt}$ are important control indicators for the preference submission timing. α_j, γ_s, and δ_t are province, track, and year fixed effects, controlling for provincial time-invariant characteristics, time-invariant differences between tracks, and

[16] The mean for raw rank difference is around six thousand; the average sum degree of envy of student rankings is 1.71 million.

[17] Over 90% of Chinese population is Han.

nationwide yearly shocks.[18] Finally, we also control for student i's percentile rank and the matching size N_{jst}, since they could also influence the likelihood and degree of justified envy. The standard errors are clustered at the provincial level.

We modify Equation 1 when looking at matching game level justified envy measures as the outcome variables:

$$\begin{aligned} Y_{jst} &= \beta_0 + \beta_1 Parallel_{jt} + \beta_2 Ex-interim_{jt} + \beta_3 Ex-post_{jt} \\ &\quad +\alpha_j + \gamma_s + \delta_t + u_{jst} \end{aligned} \tag{2}$$

While it is important to find out the average effects on justified envy and student outcomes, it is also worth investigating whether the effects persist over time. We adopt an event study model and add a set of dummy variables to indicate the number of years before or after the parallel mechanism reform. The regression model is shown in Equation 3.

$$\begin{aligned} Y_{ijst} &= \beta_0 + \sum_{k=-9}^{+9} \beta_1^k D_{jt}^k + \beta_2 Ex-interim_{jt} + \beta_3 Ex-post_{jt} \\ &\quad +\alpha_j + \gamma_s + \delta_t + u_{ijst} \end{aligned} \tag{3}$$

D_{ijst}^k is a dummy variable set to 1 if province j experienced the reform in year $t-m$ and 0 otherwise. If province j shifted from the IA to parallel mechanism k years before year t, k is positive. Otherwise, k is negative. The omitted year is $m=0$, that is, the last year province j used the IA mechanism.

Finally, we also look into whether the effect sizes are larger for provinces that adopted a larger number of parallel choices in their first choice band. We replace the dummy variable for parallel mechanism with a set of dummy indicators for different numbers of parallel choices adopted.

$$\begin{aligned} Y_{ijst} &= \beta_0 + \beta_1 No.Choice = 3_{jt} + \beta_2 No.Choice = 4_{jt} + \beta_3 No.Choice = 5_{jt} \\ &\quad +\beta_4 No.Choice = 6_{jt} + \beta_5 Ex-interim_{jt} + \beta_6 Ex-post_{jt} \\ &\quad +\beta_7 rank_i + \beta_8 N_{jst} + \alpha_j + \gamma_s + \delta_t + u_{ijst} \end{aligned} \tag{4}$$

[18]We check for the weights in this two-way fixed effects specification using *twowayfeweights* Stata package developed by de Chaisemartin and d'Haultfoeuille (2020). Only seven out of 67 ATTs receive a negative weight, which is a small proportion.

4.2 Ningxia Data Analyses

For student outcome measures using Ningxia data, we adapt Equation 1 to adjust for the fact that we only have one province and that there was no submission timing reform during the sample period for Ningxia. We control for a vector of individual characteristics X_i, including gender, household registration status (rural or urban), and ethnicity (Han or minority).

$$Y_{ist} = \beta_0 + \beta_1 Parallel_t + \beta_2 X_{ist} + \gamma_s + u_{ist} \quad (5)$$

Since we are comparing the pre- and post-reform outcomes in one province only, it is not a difference-in-differences model. Two main concerns with this analysis are external validity and contemporaneous policy changes other than mechanism change that may also influence outcomes. To understand how generalizable Ningxia's results on null admission and retaking are to the national level, we conduct two comparisons between Ningxia and the national averages. First, we run the aforementioned analysis on justified envy for Ningxia only and compare the results with the main results using the entire sample. As shown in Table A2, the relative effect sizes for Ningxia are smaller than national average. The second comparison is on the overall retake percentage. In 2012, the national average of retaking percentage is 14.4%, lower than 22.2% in Ningxia. The retaking percentages are generally lower in more affluent provinces.[19] We do not have a clear prior on whether the effect of changing mechanism on null admission and retaking at the national level should be larger or smaller than the results we find in Ningxia. There is no clear reason why we may expect the effect to only exist in Ningxia and not in other provinces. As for the second concern on contemporaneous confounding policy, we are not aware of any in Ningxia during this period.

5 Results

5.1 Effects on Justified Envy Measures

5.1.1 Main Results

Table 5 presents the main results from the generalized difference-in-differences regressions specified by Equation 1 on various measures of justified envy. The first column shows the linear probability model coefficients on the likelihood of justifiably envying anyone. Changing from IA to parallel reduced the likelihood of justified envy by 4.6 percentage points, or 5.7% of sample mean. The reform also reduced the number of students one justifiably envies by 208, which is 58% of sample mean. Summing up the number of students one justifiably envy

[19] See Figure A3 for a scatter plot on provincial GDP and retake percentage in 2012.

for all the students, we arrive at the number of blocking pairs for a matching. This provides us with an empirical counterpart to compare with the experimental studies.

Columns (3) through (8) show results for degrees of justified envy in three dimensions: how much better of a university the envied students got admitted to, how much lower the envied students ranked in their CEE scores, and what is the combined degree of envy for both dimensions. The parallel reform led to a decrease of 0.494 university rank groups in the maximum degree of envy and a decrease of 303 university rank groups in total. For student ranks, the lowest-ranked student one justifiably envies is ranked 1,920 closer to oneself after the parallel mechanism change. The total improvement in student ranking degree of envy adds up to close to one million. Despite the raw effect sizes being quite different along these two dimensions, the relative effect sizes are very similar: around 21-28% of the mean for maximum degrees of envy and around 63-71% of the mean for sum degrees of envy. Finally, the effect sizes for combined degrees of envy are similar to those along the university rank group dimension.

Next, we collapse the data to the matching game level and run analysis at the game level. Table 6 presents the results, in which we can see that the relative sizes are similar to the results at the individual level. As mentioned earlier in subsubsection 3.1.1, we also construct justified envy measures according to alternative grouping of universities. Results are strongly robust and are presented in Table 7.

Since some students may be indifferent between the top university in RG_2 versus the lowest ranked university in RG_1, we used a stricter definition of justified envy. In addition to the normal conditions of the envied student ranked lower but assigned to a university in a higher rank group, the percentile rank of that university and the admitted university has to be greater than 10%. Here, we test whether results would differ if we do not impose this condition or loosen this condition. Table 8 presents the results using data constructed without this condition, or with a similar condition but only requiring 5% university rank difference. Overall, results are similar with the main result in Table 5 and consistently support our hypothesis: compared with IA, the parallel mechanism significantly improved matching stability on the intensive margin by reducing justified envy.

5.1.2 Effects Over Time: Event Study Results

Eight graphs in Figure 4 present the event study results from Equation 3 on eight justified envy measures.[20] Each figure plots the point estimates and the 95% confidence intervals of the year dummies. The outcomes from left to right, top to down are explained as follows.

[20]Figure A1 shows the event study results with outcome measures normalized to the matching game level. Results are very similar.

The first two outcomes are whether a student has any justified envy or not and the total number of students a student justifiably envies. The second row presents degree of envy in the dimension of university rank group. Maximum degree of envy (RG) equals 0 if no justified envy and equals own rank group admitted minus the best rank group for anyone who has a lower test score. Sum degree of envy (RG) equals 0 if no justified envy and equals the sum of degree of envy for all the envied students, i.e. adding up all the differences between envied student's rank group admitted and own rank group admitted. The third row presents degrees of justified envy in the student rankings dimension, which are similar to the previous two.

The graphs show that there are no pre-trends prior to the parallel reform.[21] Overall, the coefficients of the first three to four years after mechanism change are similar to the average effects found in the difference-in-differences specification. Starting from the fifth year post-reform, the effect sizes generally increase and, in some cases, continuously increase, such as for degree of envy on the university rank group admitted. Note that these longer-term effect estimates are contributed by a few early reformers, limiting our ability to make definitive claims on whether effect strengthens over time. Nevertheless, these results support that the effects of the mechanism change persist over time.

5.1.3 Effects and Parallel Choice Bandwidth

As shown in Chen and Kesten (2017), stability should improve as the number of choices in each parallel choice band increases. To explore the relationship between the effect sizes and parallel choice bandwidth, we replace the parallel dummy in Equation 1 with several dummy varaibles, each indicating for a specific number of parallel choices in the first parallel choice band. Since provinces adopted three to six parallel choices in the first parallel choice band in our sample period, we compare the matching outcomes in these province-years with those that used IA for their college admissions. Table 9 shows the results. Surprisingly, we find that matching games adopting six parallel choices in the first parallel choice band experienced weaker effects for several justified envy outcomes when compared to those adopting five parallel choices. In particular, changing from IA to parallel with five choices in the first choice band decreases the percentage of students with any justified envy by 11.9%, but the effect size for six choices is only 3.5% and is not statistically significant. Six parallel choices also produce less reduction in maximum degree of envy on both university quality dimension and student ranking dimension. However, adopting parallel mechanism with six parallel

[21]The only significant coefficient out of all eight graphs is the dummy variable for five years prior to the reform with the outcome variable being the sum of degree of envy in student rankings. We test for a linear time trend in that setting and do not find statistical significance.

choices produces similar or larger reduction in the number of students one justifiably envies and the sum degree of envy on both dimensions. The provinces that changed from zero to six choices include Henan, Hainan, Tibet, and Chongqing, some of which are very small in their matching game size while others have a larger than average matching game size.

In the individual level results, even though we control for province fixed effects as well as the number of students in the matching game, they may still not fully account for the fact that effect sizes could vary in a nonlinear way with matching game sizes. Table 10 presents the results at the matching game level, which are overall consistent with the individual level results, except that effect sizes across choice bandwidths are closer to each other in the game level results. That being said, we still observe that the effect sizes for some of these measures do not increase linearly, not even monotonically, with the choice bandwidth. Chen and Kesten (2017) show theoretically that a wider choice band should result in better matching stability, as defined on the extensive margin: a parallel mechanism $\phi^{e=6}$ with 6 parallel choices is more stable than a parallel mechanism $\phi^{e=3}$ with 3 parallel choices. Given any matching game, if $\phi^{e=6}$ produces an unstable matching then so does $\phi^{e=3}$; if $\phi^{e=6}$ produces a stable matching, $\phi^{e=3}$ may or may not produce a stable matching. Here we explore on the intensive margin with the number of blocking pairs, which does not perfectly correspond to the binary comparison. In large matching games in real life, it is unlikely that a matching outcome can be stable, thus the intensive margin comparison is important. Further research is needed to understand and reconciliate this contrast.

5.2 Distribution of Justified Envy across Student Performance

In Figure 6, we collapse data to the percentile-mechanism level to plot the distribution of justified envy measures across student performance and compare across two mechanisms. In all eight figures, we see two hump-shaped curves, showing that those with highest and lowest CEE scores are unlikely to have justified envy. It is easy to understand it for low-performing students, that they are less likely to encounter another student with a lower test score than themselves. The more interesting part is that the high achieving students have a lot of possibility to envy others if the assignment is unfair, however, they fair better than the medium-ranked students. This is likely because the highest ranked students face less congestion and less uncertainty: they know more precisely their chances of being admitted to the very top colleges. For the medium-ranked students, they have a harder time to gauge their chance of getting the last few spots left in the top colleges over the risk of not being admitted to any of the parallel choices or the only choice in the IA mechanism. This is related to how many universities may be horizontally similar to one another, which creates congestion.

An anecdotal phenomena called ”big-small-years” portrays the struggle some students face among the mid-ranked universities. It describes the fact that some universities have surprisingly high or low cutoff scores. Quite a few universities are of similar rankings; when many students choose one particular university, it results in a high admission cutoff, “a big year” for that university. Next year, students may want to avoid this university and all choose other similarly ranked universities, resulting in a low demand and a low cutoff, “a small year” for that university.

If congestion explains the distribution of justified envy, we should expect to see higher variations in admission cutoffs for the mid-ranked universities while lower variations for the highest and lowest ranked universities in our data. Indeed, we find such relationship in Figure 7, which plots the relationship between the selectivity on the X-axis using the average cutoff scores and the variability of the cutoff scores on the Y-axis using university-level data. Each dot on the scatter plots is one university's average and standard deviation of cutoff scores between 2005 and 2011. The left panel shows the overall data, but the low standard deviation in the right tail may be caused by low-ranked universities coming in and out of the sample. Reassuringly, when we restrict to a balanced panel of universities in the right panel, the aforementioned pattern still holds.

Another clear pattern we can see from all eight graphs is that IA produces a higher level of justified envy for most measures at all levels.[22] For the binary measure of any justified envy and the maximum degree of envy along the student rank dimension, changing from IA to parallel benefits mostly the top 30% of the Tier I students. For maximum degree of envy along the university rank group dimension and the diagonal dimension, parallel mechanism improves for most students evenly. For the number of justifiably envied students and the sum degrees of envy, both mechanisms produce a normal distribution and the improvement also follows a roughly normal distribution. This suggests that the improvement on the intensive margin falls on those who had higher initial justified envy.

5.3 Effect on no-admission and retaking using Ningxia data

In addition to the justified envy measures, we also use a richer data set from Ningxia to investigate whether the parallel reform improves student admission outcomes by reducing the number of students who failed to be assigned to any school and the number of students who decided to repeat. Table 11 presents the marginal effects from a Probit model specified in Equation 5. In columns (1) and (2), the outcome variable is whether a student ended up not being assigned from any university. In columns (3) and (4), the outcome variable is

[22] We estimate statistical significance of the difference and plot the coefficients in Figure A2. Most of the time, the gaps between IA and Parallel in Figure 6 are statistically significant.

whether a student repeated a year and retook the CEEs in the following year. As we need data from two consecutive years to identify the retakers, the sample sizes in columns (3) and (4) are smaller, including only the students who first took the CEEs in 2002 and 2009.

Columns (1) and (3) present results using the full sample. Compared with IA, the parallel mechanism reduced the probability of not being assigned to by any university by 5.5% and the probability of retaking by 1.8%. These effect sizes correspond to 15 and 24.4 percentage points of the sample averages. Since we restricted our analyses to Tier I universities with the national data due to a variety of reasons described in subsection 2.2, we restrict the sample to high-achieving students with CEE scores above the Tier I university admission cutoffs to look at results for a similar population here. Results in Columns (2) and (4) suggest that adopting the parallel mechanism reduced the probability of null-admission by 0.3% and the probability of retaking by 0.5%. Although the effect sizes are much smaller than the overall sample, they are still substantial when compared with the sample mean. High-scoring students on average have much lower probabilities of not being assigned and repeating. Overall, results suggest that the parallel mechanism significantly improves student outcomes.

Since there is no reason to believe that students would predict differential improvement in their test score or face different costs of repeating their senior year in high school, the main explanatory factor for the change is the difference in student satisfaction with their admission outcome. The reduction in this revealed preference measure of retaking implies that there is an improvement in student outcomes, at least in terms of the incidences of extremely unsatisfied outcomes.

6 Robustness Checks

We conduct several robustness checks to support our findings. We first show that our results hold when we take into account other reform details or changes during the same period, which may affect our results. We then show that our results hold even when dropping the most and least selective majors from each university.

Controlling for dosage effect of the ex-post submission reform

As we have discussed in the policy background section (see Table 1), during the parallel reform process, some provinces also switched the timing of preference submission from ex-ante and ex-interim to ex-post. Provinces differ in the steps to adopt this reform. Even though we have already control for the type of preference submission timing in our model, the parallel coefficient maybe capturing the dosage effect of the ex-post submission reform. To taking into account the cumulative effect of ex-post submission reform, we add a control variable in our regression to control for how many years since each province adopted the ex-post preference submission reform. We report the results in Panel A in Table 12. Adding dosage control does not affect our main results, still supporting that the parallel reform reduces justified envy measures.

Controlling for partial parallel reform

At the beginning of the reform period, some provinces adopted a partial parallel matching mechanism. For example, they continued to use the IA mechanism for the first choice for Tier I admission while allowing for a parallel mechanism for the second and third choices of Tier I, or for all the choices of Tier II admission. In our main analysis, we only consider the matching mechanism for the first choice of in Tier I admission, since the first choice is the most important for students who passed the admission threshold for Tier I universities. However, it is not impossible that students' college application behaviors might have changed because of the parallel mechanism reforms. To address this concern, we add a binary control variable in our regression to indicate whether the province adopted a partial parallel mechanism. Panel B in Table 12 show that results remain robust.

Tier 0 and Early Admissions

In China, before the admission process of Tier I universities, military programs, police training programs, sports-related programs, and government-funded teacher training programs conduct their admission processes. Although only a very small percentage of high-achieving students choose to participate in the Tier 0 admission, the matching mechanism may influence the matching outcomes in Tier I admission. We collect information on how each province conducted their Tier 0 admission during the period from 2005 to 2011. Anhui Province was the only province that experienced changes in Tier 0 matching mechanism, switching from the IA mechanism to parallel mechanism in either 2008 or 2009.[23] We rerun our main regressions without Anhui province and show the results in Panel C of Table 12. They remain similar to the main results.

Drop the most and least selective majors from each university

One may worry that the trade-off between major and university could change the meaning of justified envy. For example, there may be a university B ranked lower than university A on average, but university B may have a very popular major that has a high cutoff score. If a student is admitted to this most popular major in university B, they may not envy another student in university A. To test how much the outlier majors could be driving our results, we drop all individuals admitted to the majors with the highest and lowest average CEE scores

[23]Jiangsu Province adopted the parallel mechanism for Tier 0 admission during the entire sample period.

from each university and then reconstruct the justified envy measures. Results are presented in Panel D of Table 12. Effect sizes for the number of students one justifiably envies and sums of degrees are similar although slightly smaller than those in the main results, and remain statistically significant at one percent level. The effect size for the percentage of students with any justified envy is 8.15%, slightly larger than the main result.

7 Conclusions

Using millions of individual-level data on Chinese college admissions, this paper provides novel empirical evidence on how matching mechanism matters for justified envy and student outcomes. We find that the parallel mechanism reduced the proportion of students with justified envy, the number of students one envies, and various measures on the degree of justified envy. One intriguing finding is that the reduction in justified envy measures does not always increase with the number of parallel choices in the first choice band. In addition, distributional analyses show that congestion in mid-ranked universities could explain the higher levels of justified envy and larger improvements from mechanism change for students ranked in the middle. Finally, data from Ningxia between 2002 and 2010 allow us to identify the students who were not admitted to any university and the retakers in the sample. Using Ningxia data, we show that switching from IA to parallel mechanism improved student outcomes by reducing the proportion of non-admissions and the proportion of retakers. Although for many retakers, repeating a year often results in a better admission outcome, this effort generates little human capital gain and is socially costly (Krishna et al., 2018).

To understand the economic value of the reduction in retakers, we conduct a simple back-of-the-envelope calculation. First, the cost for each retaker is estimated to be at around 77 to 103 thousand USD, without taking into account the interest rate.[24] Using the total number of retakers in 2012 and extrapolating the relative effect of the mechanism change, we estimate that the mechanism change reduced the number of retakers by around 81,000 per year. Multiplying the per person cost to retake by the total reduction in retakers yields around 623 to 831 million USD.[25]

We provide the first estimates for the effects of matching mechanisms on justified envy and the number of blocking pairs in a real world setting, adding evidence for the discussions on the external validity of experimental studies. In particular, Chen and Kesten (2019) find

[24] This cost consists of direct cost, which ranges between 2,000 to 48,000 RMB for students with different test scores (Larmer, 2014), and indirect cost, which is taken from the average annual earnings for a college graduate of 4,317 *12 = 51,804 RMB (Mycos, 2018).

[25] If we use an annual return of 6% and assume the age of college entrance to be 18 and the age of retirement to be 60, the lifetime savings for each cohort is between 5.8 to 8.2 trillion USD.

in the experimental setting that the PA mechanism reduces the number of blocking pairs by 50% compared with IA in the four-school environment, and by around 18% in the six-school environment. The effect sizes in our study are around 63% reduction (or 57% if we use the matching game-level result) for the sum degree of justified envy along the university rank group dimension, which corresponds to the number of blocking pairs. This is slightly larger than the experimental findings, providing some real world estimates for comparisons.

The paper has several limitations. Although we document the improvement in justified envy and extreme student outcomes, we cannot disentangle between the direct improvement from matching mechanism and the indirect improvement from students changing their preference submissions. Due to lack of data on student true ability, we also cannot speak to the ex-ante fairness. Future studies with richer data on students' list of choices before and after the mechanism change, along with measures of students' true ability would be able to answer those important questions. To our knowledge, such data does not exist in a large scale at the national level, thus these questions may be better explored in future studies combining survey data and administrative data at a local level or in a laboratory setting.

Acknowledgment

We would like to thank Yan Chen, Jie Zheng, and seminar participants at Colgate University, Jinan University, and SUNY Buffalo for helpful comments. We thank Shuai Zhao and Chenggang Hu for their research assistance.

References

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

_ **, Joshua D Angrist, Susan M Dynarski, Thomas J Kane, and Parag A Pathak**, "Accountability and flexibility in public schools: Evidence from Boston's charters and pilots," *Quarterly Journal of Economics*, 2011, *126* (2), 699–748.

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

_ **, Parag A Pathak, Alvin E Roth, and Tayfun Sönmez**, "The Boston public school match," *American Economic Review*, 2005, *95* (2), 368–371.

_ **,** _ **, and** _ , "Strategy-proofness versus efficiency in matching with indifferences: Redesigning the NYC high school match," *American Economic Review*, 2009, *99* (5), 1954–78.

__ , Yeon-Koo Che, Parag A Pathak, Alvin E Roth, and Olivier Tercieux, "Minimizing justified envy in school choice: The design of New Orleans' OneApp," Technical Report, National Bureau of Economic Research 2017.

Agarwal, Nikhil and Paulo Somaini, "Demand Analysis using Strategic Reports: An application to a school choice mechanism," *Econometrica*, 2018, *86* (2), 391–444.

Akerlof, George, "The economics of caste and of the rat race and other woeful tales," *The Quarterly Journal of Economics*, 1976, pp. 599–617.

Basteck, Christian and Marco Mantovani, "Cognitive ability and games of school choice," *Games and Economic Behavior*, 2018, *109*, 156–183.

Bo, Shiyu, Jing Liu, Ji-Liang Shiu, Yan Song, and Sen Zhou, "Admission mechanisms and the mismatch between colleges and students: Evidence from a large administrative dataset from China," *Economics of Education Review*, 2019, *68*, 27–37.

Calsamglia, Caterina, Chao Fu, and Maia Güell, "Structural estimation of a model of school choices: The Boston mechanism vs. its alternatives," *Journal of Political Economy*, 2020, *238* (2), 642–680.

Calsamiglia, Caterina, Guillaume Haeringer, and Flip Klijn, "Constrained school choice: An experimental study," *American Economic Review*, 2010, *100* (4), 1860–74.

Cao, Yuan, "Centralized assignment mechanisms and assortative matching: Evidence from Chinese universities," *Journal of Economic Behavior & Organization*, 2020, *177*, 255–276.

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

__ and __ , "Chinese college admissions and school choice reforms: An experimental study," *Games and Economic Behavior*, 2019, *115*, 83–100.

__ and Tayfun Sönmez, "School choice: an experimental study," *Journal of Economic theory*, 2006, *127* (1), 202–231.

__ , Ming Jiang, and Onur Kesten, "An empirical evaluation of Chinese college admissions reforms through a natural experiment," *Proceedings of the National Academy of Sciences*, 2020.

__ , __ , __ , Stéphane Robin, and Min Zhu, "Matching in the large: An experimental study," *Games and Economic Behavior*, 2018, *110*, 295–317.

__ , Yingzhi Liang, Tayfun Sönmez et al., "School choice under complete information: An experimental study," *Journal of Mechanism and Institution Design*, 2016, *1* (1), 45–82.

Cullen, Julie Berry, Brian A Jacob, and Steven Levitt, "The effect of school choice on participants: Evidence from randomized lotteries," *Econometrica*, 2006, *74* (5), 1191–1230.

de Chaisemartin, Clément and Xavier d'Haultfoeuille, "Two-way fixed effects estimators with heterogeneous treatment effects," *American Economic Review*, September 2020, *110* (9), 2964–2996.

Deming, David J, Justine S Hastings, Thomas J Kane, and Douglas O Staiger, "School choice, school quality, and postsecondary attainment," *American Economic Review*, 2014, *104* (3), 991–1013.

Dillon, Eleanor Wiske and Jeffrey Andrew Smith, "The consequences of academic match between students and colleges," *Journal of Human Resources*, 2020, *55* (3), 767–808.

Dobbie, Will and Roland G Fryer Jr, "Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone," *American Economic Journal: Applied Economics*, 2011, *3* (3), 158–87.

Ergin, Haluk and Tayfun Sönmez, "Games of school choice under the Boston mechanism," *Journal of public Economics*, 2006, *90* (1-2), 215–237.

Gale, David and Lloyd S Shapley, "College admissions and the stability of marriage," *American Mathematical Monthly*, 1962, *69* (1), 9–15.

Ha, Wei, Le Kang, and Yang Song, "College matching mechanisms and matching stability: Evidence from a natural experiment in China," *Journal of Economic Behavior & Organization*, 2020, *175*, 206–226.

Hoekstra, Mark, "The effect of attending the flagship state university on earnings: A discontinuity-based approach," *The Review of Economics and Statistics*, 2009, *91* (4), 717–724.

Hou, Dingkai, Meijiao Zhang, and Xiaona Li, "The effect of the parallel mechanism in Shanghai," *Research in Education Development*, 2009, *7*, 27–31.

Kapor, Adam, Christopher A Neilson, and Seth D Zimmerman, "Heterogeneous beliefs and school choice mechanisms," Technical Report, National Bureau of Economic Research 2018.

Krishna, Kala, Sergey Lychagin, and Veronica Frisancho, "Retaking in High Stakes Exams: Is Less More?," *International Economic Review*, 2018, *59* (2), 449–477.

Larmer, Brook, "Inside a Chinese test-prep factory," *New York Times*, 2014.

Levitt, Steven D and John A List, "What do laboratory experiments measuring social preferences reveal about the real world?," *Journal of Economic perspectives*, 2007, *21* (2), 153–174.

Li, Bingbing, Stephanie Wang, and Xiaohan Zhong, "Timing is everything? An experimental investigation of incomplete information in centralized matching mechanisms," *Manuscript, Univ. Pittsburgh. http://www. pitt. edu/˜ swwang/papers/Matching. pdf*, 2016.

Li, Feng, G Li, and Xiao ling Yang, "Research on application timing and admission mechanism in national college entrance exam," *Educational Research*, 2010, *10*, 53–59.

Li, Hongbin, Lingsheng Meng, Xinzheng Shi, and Binzhen Wu, "Does attending elite colleges pay in China?," *Journal of Comparative Economics*, 2012, *40* (1), 78–88.

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

_ , _ , and _ , "Ex-ante fairness in the Boston and serial dictatorship mechanisms under pre-exam and post-exam preference submission," *Games and Economic Behavior*, 2017, *101*, 98–120.

Mycos, Consulting, "Annual Report on Labor Market Outcomes of Chinese College Graduates: 2018," 2018.

Nie, Haifeng, "A game-theoretical analysis of China's college admission mechanism," *China Economic Quarterly*, 2007, *6* (3), 899–916.

Pathak, Parag A and Tayfun Sönmez, "Leveling the playing field: Sincere and sophisticated players in the Boston mechanism," *American Economic Review*, 2008, *98* (4), 1636–52.

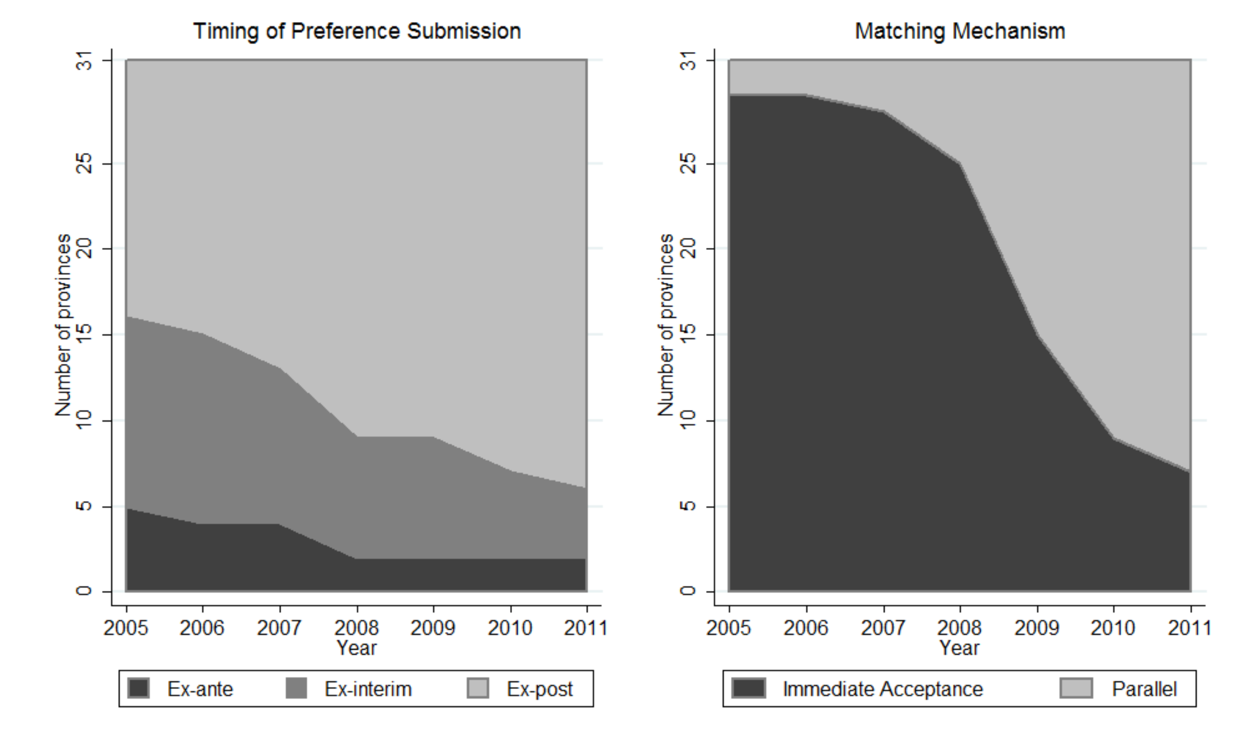
Shen, B, C Wang, and Z Lu, "A successful attempt to improve fairness in college admissions — A first victory of parallel mechanism reform in Admissions to Shanghai universities," *Chinese Higher Education*, 2008, *19*, 44–46.

Wu, Binzhen and Xiaohan Zhong, "Matching mechanisms and matching quality: Evidence from a top university in China," *Games and Economic Behavior*, 2014, *84*, 196–215.

Zhong, Xiaohan, Na Cheng, and Yunfan He, "Where have all the flowers gone — An analysis of the national college entrance exam," *China Economic Quarterly*, 2004, *3*, 763–778.

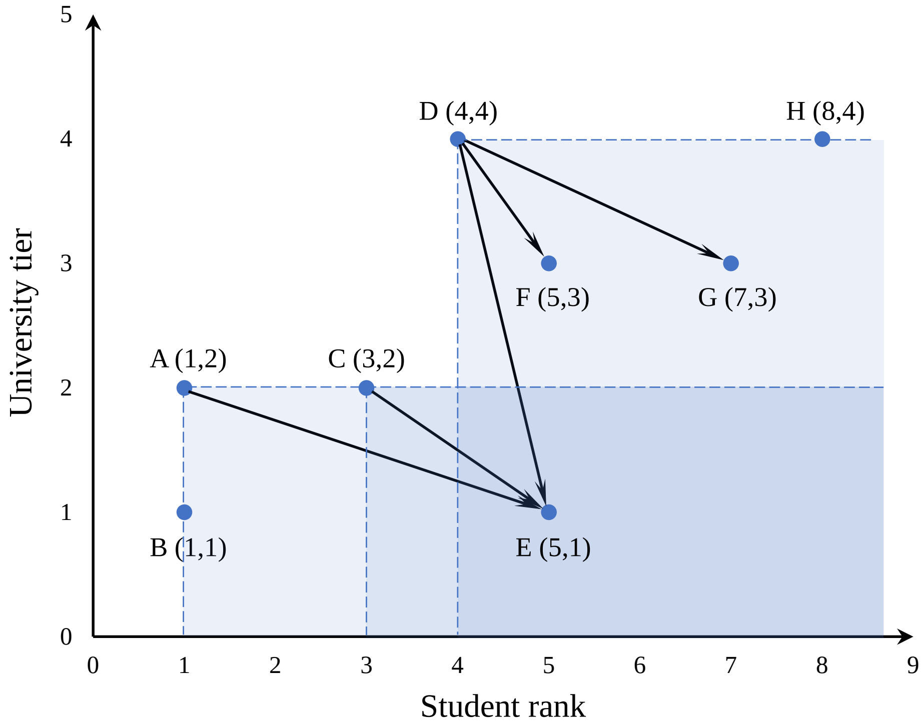
Zhu, Min, "College admissions in China: A mechanism design perspective," *China Economic Review*, 2014, *30*, 618–631.

Figure 1: Provincial Variations in the Timing of College Admissions Reform



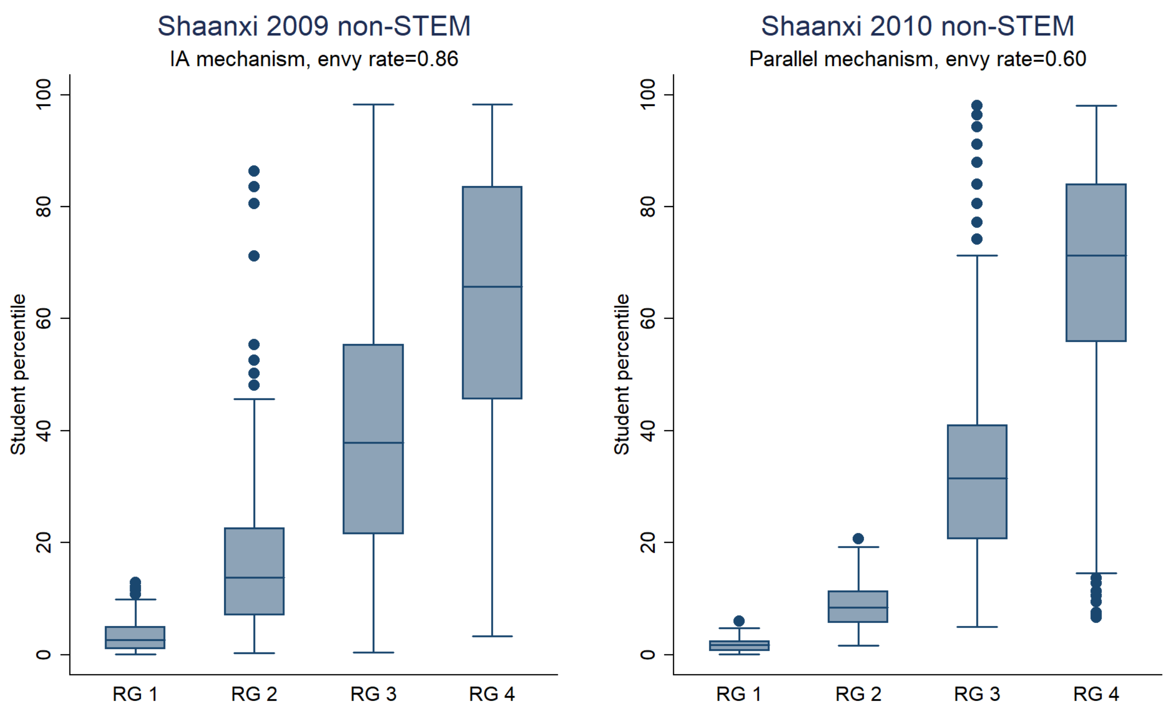
Notes: The left panel shows the number of provinces that adopted choice submission after knowing test scores (Ex-post), choice submission after the exam without knowing test scores (Ex-interim), and choice submission before the exam (Ex-ante) across years. The right panel shows the number of provinces that used the Boston or IA mechanism versus the PA mechanism between 2005 and 2011.

Figure 2: Illustrative Example for Justified Envy Measures



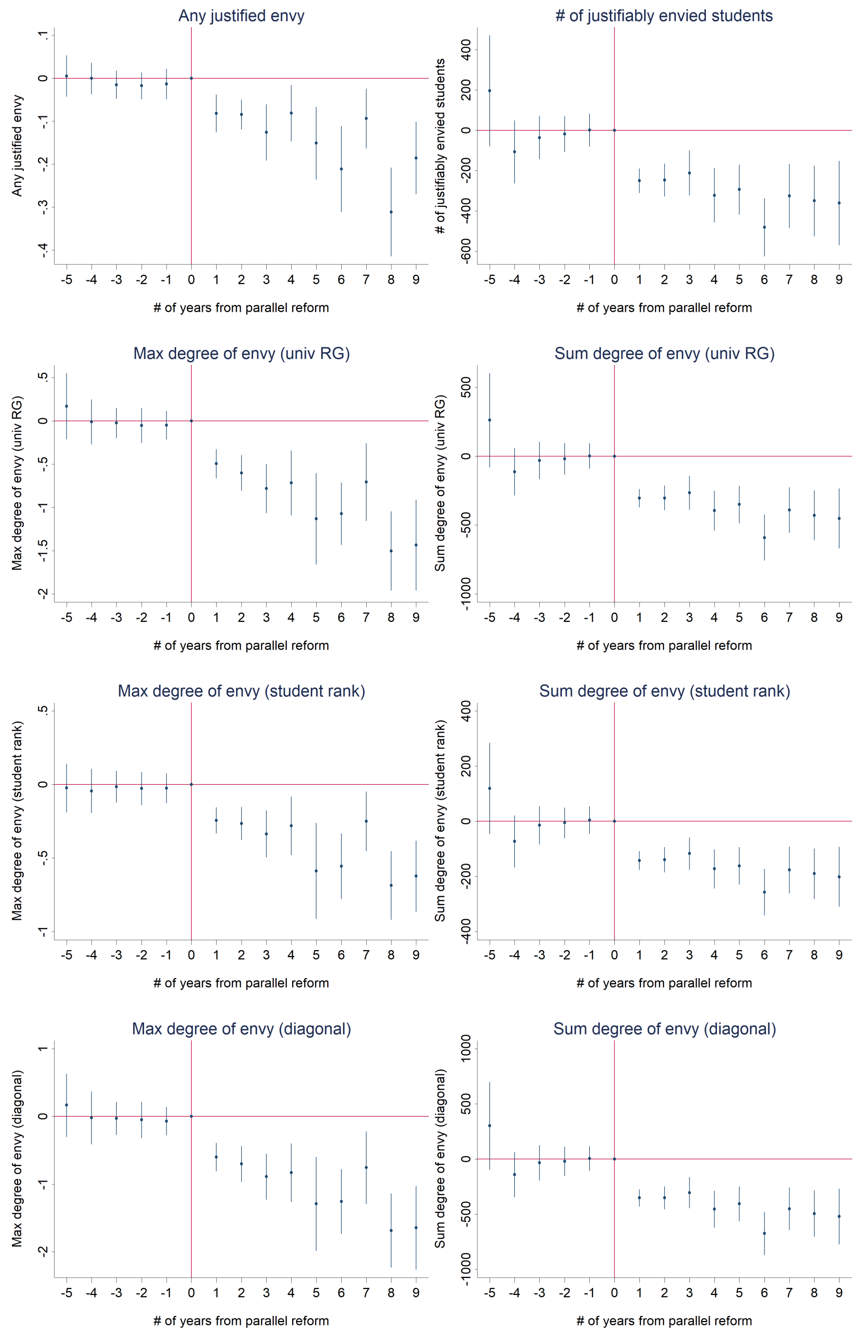
Notes: Smaller numbers mean higher ranks. The shaded areas indicate the areas of possible justified envy. The arrows point from the students with justified envy to the student(s) they justifiably envy.

Figure 3: Illustrative Example of Shaanxi Non-STEM Track: Distribution of student score percentile under IA and Parallel mechanism



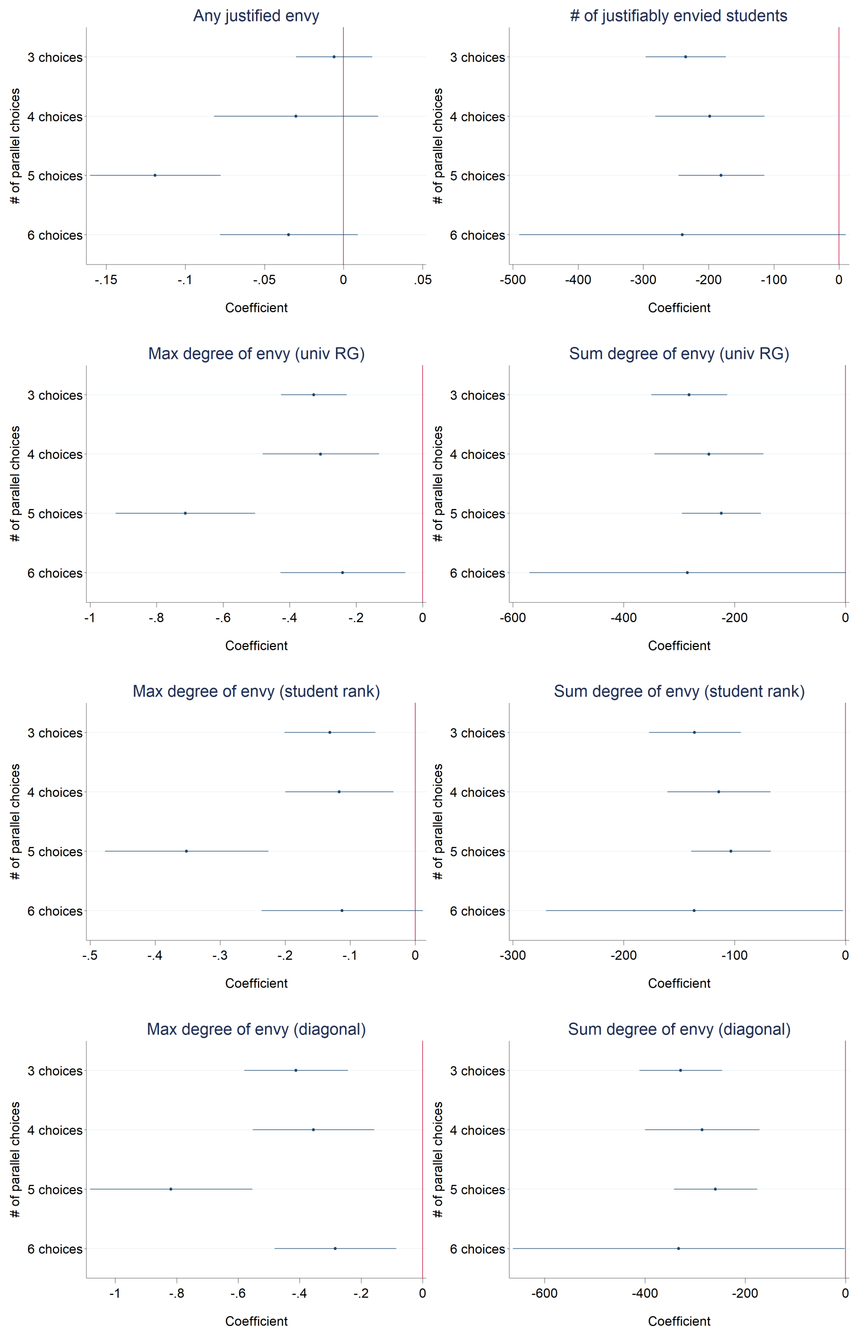
Notes: These two figures illustrate the standard box plots of student CEE percentiles for each rank group of universities in two matching games, one for Shaanxi 2009 Non-STEM track, and the other for Shaanxi 2010 Non-STEM track. The upper and lower borders of each box show the 25th and 75th percentiles of student admitted to that rank group, the horizontal line within the boxes show the medians. The outer borders show the adjacent values (25th percentile -1.5 interquartile range, 75th percentile + 1.5 interquartile range).

Figure 4: Event Study Results at the Individual Level



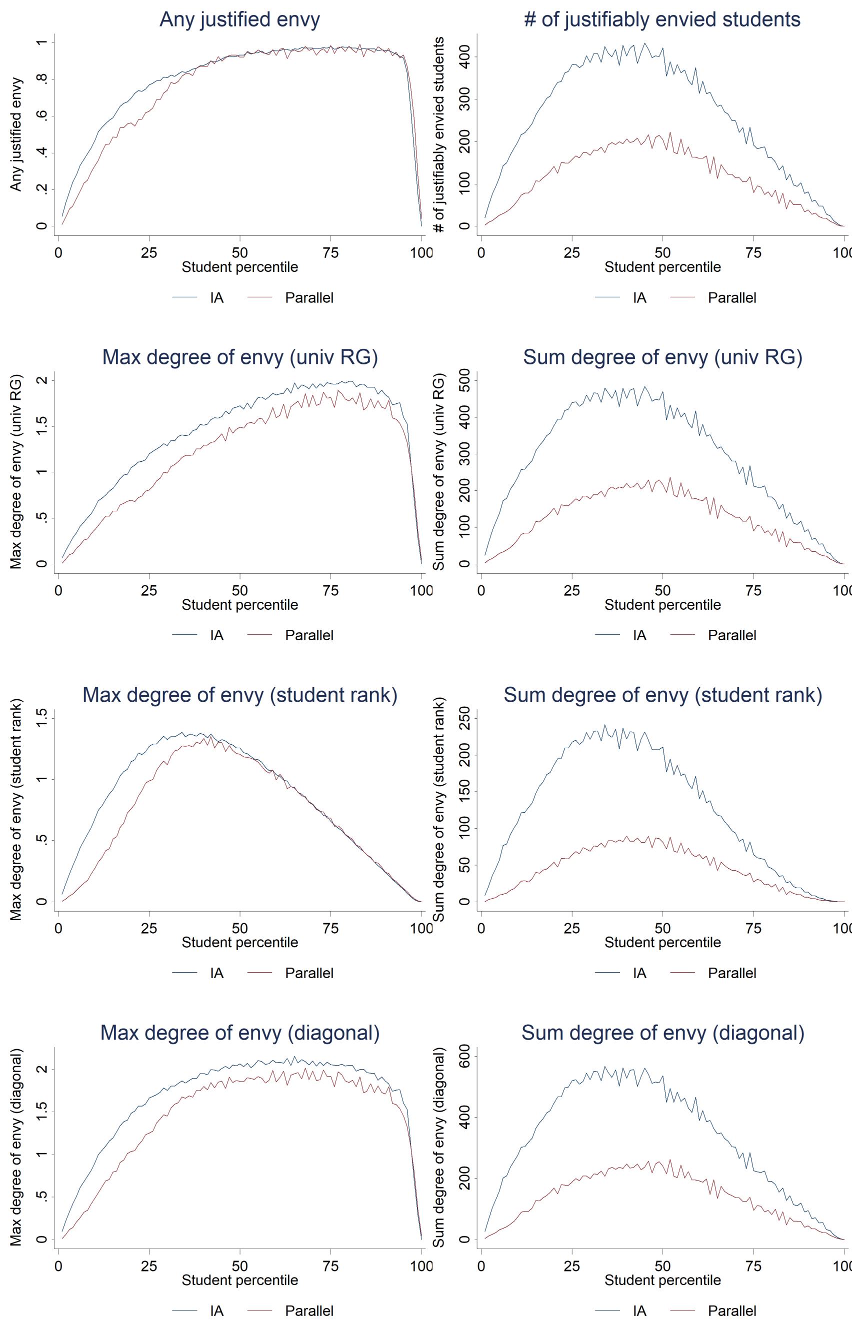
Notes: These figures present the event study coefficients of Equation 3. The outcomes from left to right, top to down are explained as follows. The first two outcomes are whether a student has any justified envy or not and the total number of students a student justifiably envies. The second row presents degree of envy in the dimension of university rank group. Maximum degree of envy (RG) equals 0 if no justified envy and equals own rank group admitted minus the best rank group for anyone who has a lower test score. Sum degree of envy (RG) equals 0 if no justified envy and equals the sum of degree of envy for all the envied students, i.e. adding up all the differences between envied student's rank group admitted and own rank group admitted. The third row presents degrees of justified envy in the student rankings dimension, which are similar to the previous two. The criteria used to assign a university into a rank group is the average admission score percentile rank that year. The graphs show that there are no pre-trends prior to the parallel reform.

Figure 5: Effect Sizes for Different Numbers of Parallel Choices



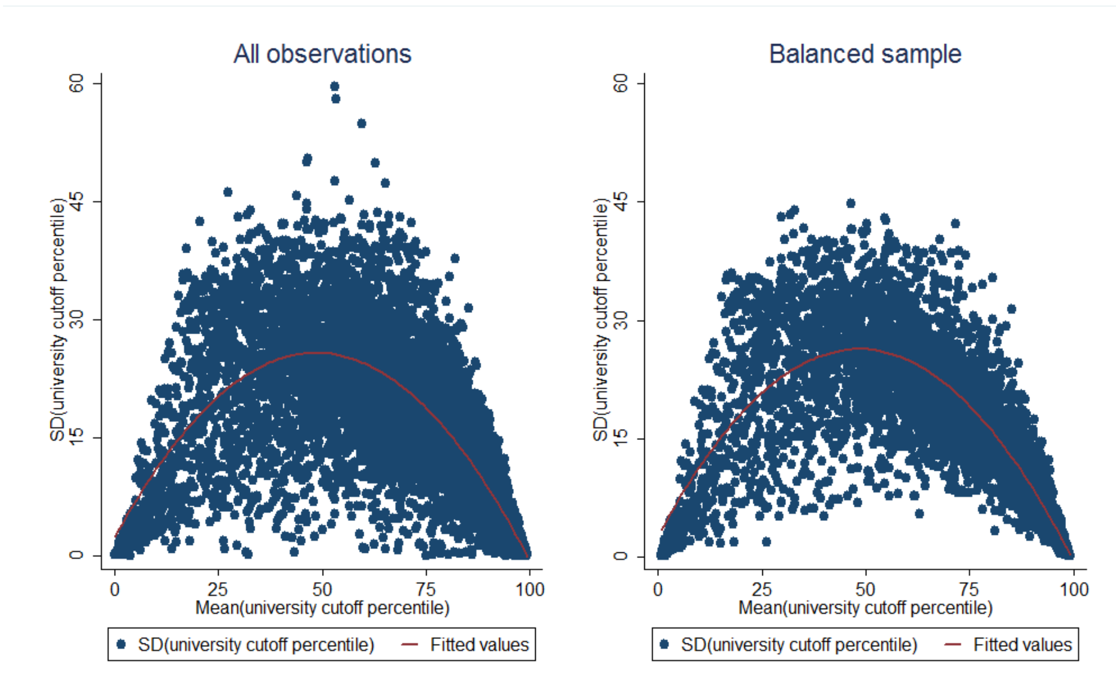
Notes: These figures present the coefficients of Equation 4. The outcomes from left to right, top to down are explained as follows. The first two outcomes are whether a student has any justified envy or not and the total number of students a student justifiably envies. The second row presents degree of envy in the dimension of university rank group. Maximum degree of envy (RG) equals 0 if no justified envy and equals own rank group admitted minus the best rank group for anyone who has a lower test score. Sum degree of envy (RG) equals 0 if no justified envy and equals the sum of degree of envy for all the envied students, i.e. adding up all the differences between envied student's rank group admitted and own rank group admitted. The third row presents degrees of justified envy in the student rankings dimension, which are similar to the previous two. The last row presents degrees of diagonal justified envy, combining two previous dimensions. The criteria used to assign a university into a rank group is the average admission score percentile rank that year.

Figure 6: Distribution of Justified Envy



Notes: These figures present the average of various justified envy measures along the CEE score distribution for matching games with IA versus parallel mechanisms. Note that in all our analyses, we limit our sample to those with a CEE score above the Tier I cutoff score. A lower percentile means a higher CEE score.

Figure 7: Distribution of Congestion at the University Level



Notes: These figures present the relationship between the standard deviation and mean of university cutoff score percentile across the sample years. The left figure shows data for all universities that admit students in Tier I admissions; the right figure shows data for universities that consistently admit students in Tier I admissions between 2005 and 2011. They show that universities that are ranked the highest and lowest among the Tier I universities have low variations of cutoff score changes across years, indicating less congestion for students whose scores fall into these ranges.

Table 1: Timing of policy reform in each province

Province	parallel	ex-interim	ex-post
Beijing	2014	-	2015
Tianjin	2010	before 2005	2011
Hebei	2009	-	1999
Shanxi	2012	before 2005	2012
Inner Mongolia	-	-	2002
Liaoning	2008	before 2005	2014
Jilin	2009	-	2008
Heilongjiang	2013	before 2005	2013
Shanghai	2008	-	2017
Jiangsu	2005	-	2003
Zhejiang	2007	-	1999
Anhui	2008	before 2005	2007
Fujian	2009	-	2005
Jiangxi	2009	before 2005	2007
Shandong	2013	-	1998
Henan	2010	before 2005	2010
Hubei	2011	-	2014
Hunan	2003	-	2001
Guangdong	2010	-	2008
Guangxi	2009	-	2004
Hainan	2009	-	2002
Chongqing	2010	-	2006
Sichuan	2009	-	2005
Guizhou	2009	before 2005	2008
Yunnan	2009	-	2004
Tibet	2010	-	1996
Shaanxi	2010	before 2005	2010
Gansu	2015	before 2005	2008
Qinghai	2018	-	1996
Ningxia	2009	-	1999
Xinjiang	2011	before 2005	2015

Table 2: Illustrative Example for Justified Envy Measures

Student ID	A	B	C	D	E	F	G	H
Student rank	1	1	3	4	5	5	7	8
University rank group admitted	2	1	2	4	1	3	3	4
Any justified envy (0/1)	1	0	1	1	0	0	0	0
Number of students one justifiably envies	1	0	1	3	0	0	0	0
Max degree of envy (university rank group)	1	0	1	3	0	0	0	0
Sum degree of envy (university rank group)	1	0	1	5	0	0	0	0
Max degree of envy (student rank)	4	0	2	3	0	0	0	0
Sum degree of envy (student rank)	4	0	2	5	0	0	0	0
Max degree of envy (diagonal)	$\sqrt{17}$	0	$\sqrt{5}$	$\sqrt{10}$	0	0	0	0
Sum degree of envy (diagonal)	$\sqrt{17}$	0	$\sqrt{5}$	$2 * \sqrt{10} + \sqrt{2}$	0	0	0	0

Table 3: Summary Statistics of National Data

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
Rank	3.73e+06	8,696	7,853	1	47,248
Percentile rank	3.73e+06	49.27	28.59	0.00205	100
Parallel mechanism	3.73e+06	0.373	0.484	0	1
Ex-ante submission	3.73e+06	0.0827	0.275	0	1
Ex-interim submission	3.73e+06	0.249	0.433	0	1
Ex-post submission	3.73e+06	0.668	0.471	0	1
STEM track	3.73e+06	0.816	0.388	0	1
Justified envy neasures					
Any justified envy (0/1)	3.730e+06	0.806	0.396	0	1
Number of justifiably envied students	3.730e+06	357.2	951.6	0	38,414
Max degree of envy (univ RG)	3.730e+06	1.467	0.959	0	3
Sum degree of envy (univ RG)	3.730e+06	402.9	1,126	0	62,815
Max degree of envy (rank)	3.730e+06	0.870	0.740	0	2.966
Sum degree of envy (rank)	3.730e+06	168.0	576.7	0	45,948
Max degree of envy (diagonal)	3.730e+06	1.742	1.032	0	4.196
Sum degree of envy (diagonal)	3.730e+06	455.9	1,327	0	82,887

Notes: This table reports summary statistics of individual-level Chinese College Entrance Exam and college admissions outcomes, after excluding those admitted to universities in Hong Kong or admitted through art, music, and sports programs. Four university rank groups (RGs) are constructed using rankings based on the average admission cutoff score percentile for the university within each matching game at the track-province-year level. More details on the definition of justified envy measures are in subsubsection 3.1.2.

Table 4: Summary Statistics of Ningxia Data

Tier ALL VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
Null admission	417,690	0.366	0.482	0	1
Retake CEE	363,043	0.224	0.417	0	1
Parallel mechanism	417,690	0.264	0.441	0	1
CEE score	417,690	390.6	87.95	12	727
Male	417,690	0.520	0.500	0	1
Hukou: urban	417,690	0.449	0.497	0	1
Ethnicity: han	417,690	0.777	0.417	0	1
STEM track	417,690	0.667	0.471	0	1

Note: This table reports summary statistics of Ningxia individual level data between 2002 to 2010. Retaking CEE in the next year has fewer observations because we need two consecutive years to identify the retakers from the data.

Table 5: Matching Mechanisms and Justified Envy

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Degree of Envy Dimension		University Rank Group		Student Rank		Diagonal	
Dependant variable	Any justified envy (0/1)	Number of justifiably envied students	Max degree	Sum Degree	Max degree	Sum envy	Max degree	Sum Degree
Parallel	-0.0461**	-208.3***	-0.415***	-254.8***	-0.182***	-119.9***	-0.489***	-295.9***
	(0.0199)	(24.18)	(0.0848)	(26.74)	(0.0447)	(14.15)	(0.101)	(31.50)
Ex-interim	-0.0803***	187.6	-0.378**	212.8	-0.192*	105.1	-0.395*	249.5
	(0.0289)	(116.1)	(0.145)	(133.1)	(0.0971)	(68.52)	(0.206)	(157.3)
Ex-post	-0.0495	-109.5*	-0.00827	-114.8*	-0.0327	-64.61*	0.0503	-138.0*
	(0.0297)	(57.15)	(0.133)	(68.60)	(0.0981)	(35.27)	(0.200)	(81.08)
Observations	3,729,727	3,729,727	3,729,727	3,729,727	3,729,727	3,729,727	3,729,727	3,729,727
R-squared	0.199	0.077	0.337	0.072	0.097	0.062	0.224	0.070
Mean	0.806	357.2	1.467	402.9	0.870	168	1.742	455.9
SD	0.396	951.6	0.959	1126	0.740	576.7	1.032	1327
effect/mean	-0.0572	-0.583	-0.283	-0.632	-0.210	-0.714	-0.281	-0.649
effect/sd	-0.116	-0.219	-0.432	-0.226	-0.247	-0.208	-0.474	-0.223

Notes: Each cell reports the coefficient of parallel reform from a separate OLS regression, including track fixed effects, province fixed effects, year fixed effects, reform timing controls, the total number of students participating in CEE in the track-province that year, and individual's score percentile. Robust standard errors clustered at the province level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Results for Justified Envy Measures Normalized to the Matching Game Level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Degree of Envy Dimension		University Rank Group		Student Rank		Diagonal	
Dependant variable	Any justified envy (0/1)	Number of justifiably envied students	Max degree	Sum Degree	Max degree	Sum envy	Max degree	Sum Degree
Parallel	-0.0713***	-0.0440***	-0.173***	-0.0333***	-0.106***	-0.0196***	-0.158***	-0.0295***
	(0.0144)	(0.00374)	(0.0260)	(0.00310)	(0.0214)	(0.00194)	(0.0264)	(0.00276)
Ex-interim	-0.0589*	0.0142*	-0.145*	0.0104*	-0.0992	0.00711*	-0.135	0.00959*
	(0.0328)	(0.00755)	(0.0828)	(0.00601)	(0.0807)	(0.00367)	(0.0953)	(0.00515)
Ex-post	-0.0407	-0.0232***	-0.0382	-0.0177***	-0.0501	-0.00874***	-0.0307	-0.0144***
	(0.0300)	(0.00412)	(0.0798)	(0.00388)	(0.0785)	(0.00217)	(0.0936)	(0.00325)
Observations	398	398	398	398	398	398	398	398
R-squared	0.531	0.796	0.603	0.779	0.529	0.736	0.559	0.769
Mean	0.864	0.0784	0.689	0.0581	0.559	0.0284	0.647	0.0489
SD	0.0973	0.0373	0.184	0.0290	0.143	0.0171	0.173	0.0253
effect/mean	-0.0825	-0.562	-0.251	-0.573	-0.189	-0.691	-0.244	-0.605
effect/sd	-0.733	-1.180	-0.943	-1.147	-0.738	-1.143	-0.914	-1.165

Notes: Each cell reports the coefficient of parallel reform from a separate OLS regression, including track fixed effects, province fixed effects, year fixed effects, reform timing controls, and the total number of students participating in CEE in the track-province that year. The outcome variables are standardized by dividing the sum of each variable at the cell level by the sum of the eight outcomes in the worst matching outcome possible for that specific cell. Robust standard errors clustered at the province level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Results using Other Rank Group Assignemnts for Universities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. 4 Tiers Based on 985/211								
Dependant variable	Any justified envy (0/1)	Number of justifiably envied students	Max degree of envy (univ RG)	Sum degree of envy (univ RG)	Max degree of envy (rank)	Sum degree of envy (rank)	Max degree of envy (diagonal)	Sum degree of envy (diagonal)
Parallel	-0.0494***	-150.7***	-0.119**	-181.3***	-0.169***	-88.64***	-0.214***	-211.4***
	(0.0175)	(20.22)	(0.0457)	(22.53)	(0.0312)	(13.19)	(0.0502)	(26.97)
Mean	0.702	272.8	1.107	336.6	0.727	130.8	1.378	376.2
SD	0.457	741.1	0.864	951.2	0.726	454.5	1.014	1098
effect/mean	-0.0704	-0.552	-0.107	-0.539	-0.233	-0.678	-0.155	-0.562
effect/sd	-0.108	-0.203	-0.137	-0.191	-0.233	-0.195	-0.211	-0.192
Panel B. 4 Tiers Based on Previous Year's School Rank								
Parallel	-0.0463**	-200.2***	-0.444***	-265.0***	-0.294***	-145.9***	-0.612***	-319.6***
	(0.0201)	(30.60)	(0.0892)	(37.12)	(0.0545)	(22.27)	(0.112)	(45.09)
Mean	0.853	404	1.723	473.9	1.114	226	2.125	553.1
SD	0.354	903.1	1	1099	0.868	615.6	1.080	1328
effect/mean	-0.0543	-0.496	-0.257	-0.559	-0.263	-0.646	-0.288	-0.578
effect/sd	-0.131	-0.222	-0.444	-0.241	-0.338	-0.237	-0.566	-0.241
Panel C. 4 Tier Based on Average Rank 2005-2011								
Parallel	-0.0555***	-210.9***	-0.488***	-274.0***	-0.225***	-134.7***	-0.597***	-321.3***
	(0.0196)	(35.39)	(0.0801)	(37.27)	(0.0453)	(18.95)	(0.0933)	(43.37)
Mean	0.826	415.2	1.580	478.9	0.946	208.7	1.902	546.8
SD	0.379	1017	0.991	1235	0.769	656.7	1.053	1469
effect/mean	-0.0671	-0.508	-0.308	-0.572	-0.238	-0.646	-0.314	-0.588
effect/sd	-0.146	-0.207	-0.492	-0.222	-0.293	-0.205	-0.567	-0.219
Panel D. 8 Tiers Based on Current Year's School Rank								
Parallel	-0.0150	-273.8***	-0.740***	-544.9***	-0.315***	-375.8***	-0.813***	-720.9***
	(0.0128)	(31.31)	(0.155)	(65.98)	(0.0890)	(45.16)	(0.177)	(85.61)
Mean	0.831	535.3	3.027	888.4	2.178	594.1	3.768	1157
SD	0.375	1241	1.889	2344	1.759	1853	2.085	3247
effect/mean	-0.0181	-0.512	-0.244	-0.613	-0.145	-0.633	-0.216	-0.623
effect/sd	-0.0401	-0.221	-0.392	-0.232	-0.179	-0.203	-0.390	-0.222
Panel E. 8 Tiers Based on Previous Year's School Rank								
Parallel	-0.0172	-293.9***	-1.044***	-639.0***	-0.509***	-496.2***	-1.254***	-879.9***
	(0.0128)	(38.11)	(0.165)	(85.76)	(0.103)	(66.48)	(0.207)	(116.1)
Mean	0.870	608.6	3.601	1071	2.703	784.5	4.604	1444
SD	0.336	1238	2.034	2438	2.026	2041	2.238	3457
effect/mean	-0.0197	-0.483	-0.290	-0.597	-0.188	-0.633	-0.272	-0.609
effect/sd	-0.0511	-0.237	-0.513	-0.262	-0.251	-0.243	-0.561	-0.255
Panel F. 8 Tiers Based on Average Rank 2005-2011								
Parallel	-0.0240*	-274.4***	-0.959***	-592.5***	-0.359***	-417.3***	-1.041***	-787.9***
	(0.0140)	(37.19)	(0.158)	(71.67)	(0.0981)	(55.88)	(0.188)	(97.96)
Mean	0.842	607.7	3.319	1055	2.315	713.3	4.120	1379
SD	0.365	1313	2.009	2603	1.814	2050	2.171	3593
effect/mean	-0.0285	-0.452	-0.289	-0.561	-0.155	-0.585	-0.253	-0.571
effect/sd	-0.0659	-0.209	-0.477	-0.228	-0.198	-0.204	-0.480	-0.219

Notes: Each cell reports the coefficient of parallel reform from a separate OLS regression, including track fixed effects, province fixed effects, year fixed effects, and reform timing controls. The first panel uses the following grouping rules: Peking and Tsinghua are the first rank group, 985 project universities are the second RG, 211 project university (non-985) are the third RG, the rest of Tier I universities are the fourth RG. The other panels use the average admitted students' CEE scores to assign university RGs; unlike the main results, which use the current year to categorize into four groups, we use the previous year or average rank to assign rank groups. We also assign universities into eight groups instead of four groups based on the following cutoff percentiles: 5%, 10%, 20%, 30%, 45%, 60%, 80%. Robust standard errors clustered at the province level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 8: Results using a Looser Justified Envy Definition:
University Percentile Rank of student j - Own University Percentile Rank > 0% or 5%

Panel A. 0% condition								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependant variable	Any justified envy (0/1)	Number of justifiably envied students	Max degree of envy (univ RG)	Sum degree of envy (univ RG)	Max degree of envy (rank)	Sum degree of envy (rank)	Max degree of envy (diagonal)	Sum degree of envy (diagonal)
Parallel	-0.0420**	-214.8***	-0.411***	-261.3***	-0.224***	-127.8***	-0.517***	-305.9***
	(0.0193)	(28.38)	(0.0823)	(29.82)	(0.0554)	(15.24)	(0.106)	(34.74)
Observations	3,729,727	3,729,727	3,729,727	3,729,727	3,729,727	3,729,727	3,729,727	3,729,727
R-squared	0.114	0.089	0.309	0.083	0.189	0.071	0.180	0.081
Mean	0.857	436.7	1.519	482.4	0.977	203.6	1.860	547.3
SD	0.350	1056	0.904	1225	0.763	641.4	0.959	1449
effect/mean	-0.0490	-0.492	-0.270	-0.542	-0.229	-0.628	-0.278	-0.559
effect/sd	-0.120	-0.203	-0.454	-0.213	-0.293	-0.199	-0.539	-0.211
Panel B. 5% condition								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependant variable	Any justified envy (0/1)	Number of justifiably envied students	Max degree of envy (univ RG)	Sum degree of envy (univ RG)	Max degree of envy (rank)	Sum degree of envy (rank)	Max degree of envy (diagonal)	Sum degree of envy (diagonal)
Parallel	-0.0423**	-216.7***	-0.411***	-263.3***	-0.207***	-127.9***	-0.505***	-307.9***
	(0.0199)	(26.33)	(0.0831)	(28.21)	(0.0521)	(14.91)	(0.105)	(33.14)
Observations	3,729,727	3,729,727	3,729,727	3,729,727	3,729,727	3,729,727	3,729,727	3,729,727
R-squared	0.129	0.084	0.315	0.078	0.161	0.068	0.191	0.076
Mean	0.849	408.7	1.510	454.4	0.950	191.3	1.833	515.2
SD	0.358	1023	0.913	1194	0.754	622	0.972	1412
effect/mean	-0.0498	-0.530	-0.272	-0.579	-0.218	-0.669	-0.275	-0.598
effect/sd	-0.118	-0.212	-0.450	-0.220	-0.275	-0.206	-0.519	-0.218

Notes: Each cell reports the coefficient of parallel reform from a separate OLS regression, including track fixed effects, province fixed effects, year fixed effects, reform timing controls, the total number of students participating in CEE in the track-province that year, and individual's score percentile. The definition of justified envy in main results is that the difference in university rank percentile has to be at least 10% for a student i to justifiably envy student j, in addition to the basic definition that i has a higher test score than j and a lower admitted university rank group than j. Here, we loosen the restriction to be 5% or 0%. Robust standard errors clustered at the province level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 9: Matching Mechanism and Justified Envy: Nonlinearity in Bandwidth

Dependant variable	(1) Any justified envy (0/1)	(2) Number of justifiably envied students	(3) Max degree of envy (univ RG)	(4) Sum degree of envy (univ RG)	(5) Max degree of envy (rank)	(6) Sum degree of envy (rank)	(7) Max degree of envy (diagonal)	(8) Sum degree of envy (diagonal)
3 Parallel Choices	-0.00604	-235.2***	-0.328***	-282.3***	-0.131***	-136.2***	-0.412***	-329.0***
	(0.0120)	(30.66)	(0.0494)	(34.13)	(0.0348)	(20.63)	(0.0840)	(41.07)
4 Parallel Choices	-0.0301	-198.1***	-0.307***	-246.7***	-0.117***	-114.4***	-0.356***	-285.9***
	(0.0259)	(41.79)	(0.0874)	(49.06)	(0.0416)	(23.18)	(0.0986)	(56.81)
5 Parallel Choices	-0.119***	-180.9***	-0.714***	-224.2***	-0.352***	-103.5***	-0.819***	-259.5***
	(0.0205)	(32.79)	(0.105)	(35.53)	(0.0626)	(17.89)	(0.131)	(41.46)
6 Parallel Choices	-0.0347	-240.2*	-0.241**	-285.4**	-0.113*	-136.5**	-0.285***	-332.6**
	(0.0217)	(124.9)	(0.0936)	(142.5)	(0.0620)	(66.80)	(0.0986)	(165.0)
Observations	3,729,727	3,729,727	3,729,727	3,729,727	3,729,727	3,729,727	3,729,727	3,729,727
R-squared	0.202	0.077	0.345	0.072	0.101	0.063	0.232	0.070
Mean	0.806	357.2	1.467	402.9	0.870	168	1.742	455.9
SD	0.396	951.6	0.959	1126	0.740	576.7	1.032	1327

Notes: Each cell reports the coefficient of parallel reform from a separate OLS regression, including track fixed effects, province fixed effects, year fixed effects, reform timing controls, the total number of students participating in CEE in the track-province that year, and individual's score percentile. Robust standard errors clustered at the province level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 10: Matching Game Level Results: Choice Bandwidth

Dependant variable	(1) Any justified envy (0/1)	(2) Number of justifiably envied students	(3) Max degree of envy (univ RG)	(4) Sum degree of envy (univ RG)	(5) Max degree of envy (rank)	(6) Sum degree of envy (rank)	(7) Max degree of envy (diagonal)	(8) Sum degree of envy (diagonal)
3 Parallel Choices	-0.0138	-0.0404***	-0.130***	-0.0305***	-0.0651***	-0.0176***	-0.120***	-0.0270***
	(0.0198)	(0.00545)	(0.0263)	(0.00431)	(0.0243)	(0.00268)	(0.0323)	(0.00372)
4 Parallel Choices	-0.0753***	-0.0515***	-0.170***	-0.0389***	-0.0979***	-0.0232***	-0.155***	-0.0347***
	(0.0197)	(0.00604)	(0.0315)	(0.00506)	(0.0247)	(0.00304)	(0.0302)	(0.00450)
5 Parallel Choices	-0.120***	-0.0356***	-0.269***	-0.0267***	-0.189***	-0.0158***	-0.246***	-0.0238***
	(0.0190)	(0.00366)	(0.0362)	(0.00286)	(0.0348)	(0.00172)	(0.0393)	(0.00245)
6 Parallel Choices	-0.0755*	-0.0402***	-0.113**	-0.0307***	-0.0713	-0.0177***	-0.0976**	-0.0268***
	(0.0426)	(0.00799)	(0.0440)	(0.00648)	(0.0501)	(0.00422)	(0.0432)	(0.00579)
Observations	398	398	398	398	398	398	398	398
R-squared	0.557	0.802	0.625	0.784	0.554	0.742	0.580	0.775
Mean	0.864	0.0784	0.689	0.0581	0.559	0.0284	0.647	0.0489
SD	0.0973	0.0373	0.184	0.0290	0.143	0.0171	0.173	0.0253

Notes: Each cell reports the coefficient of parallel reform from a separate OLS regression, including track fixed effects, province fixed effects, year fixed effects, reform timing controls, the total number of students participating in CEE in the track-province that year, and individual's score percentile. Robust standard errors clustered at the province level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 11: Effects on Student Welfare using Ningxia Data

VARIABLES	(1) Null-admission	(2) Null-admission	(3) Retake	(4) Retake
parallel	-0.055***	-0.003**	-0.018***	-0.005***
	(0.002)	(0.001)	(0.002)	(0.002)
male	0.047***	0.002*	0.016***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)
urban hukou	-0.129***	0.002	-0.119***	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)
han ethnicity	0.050***	-0.002*	0.066***	0.004***
	(0.002)	(0.001)	(0.002)	(0.001)
STEM	-0.018***	0.002	0.019***	0.005**
	(0.002)	(0.002)	(0.001)	(0.002)
Observations	417,690	43,794	363,043	36,399
Mean	0.366	0.0114	0.224	0.0140
SD	0.482	0.106	0.417	0.117
effect/mean	-0.150	-0.244	-0.0791	-0.363
effect/sd	-0.114	-0.0262	-0.0425	-0.0431

Notes: This table reports the marginal effects from probit regressions using Ningxia data from 2002 to 2010. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

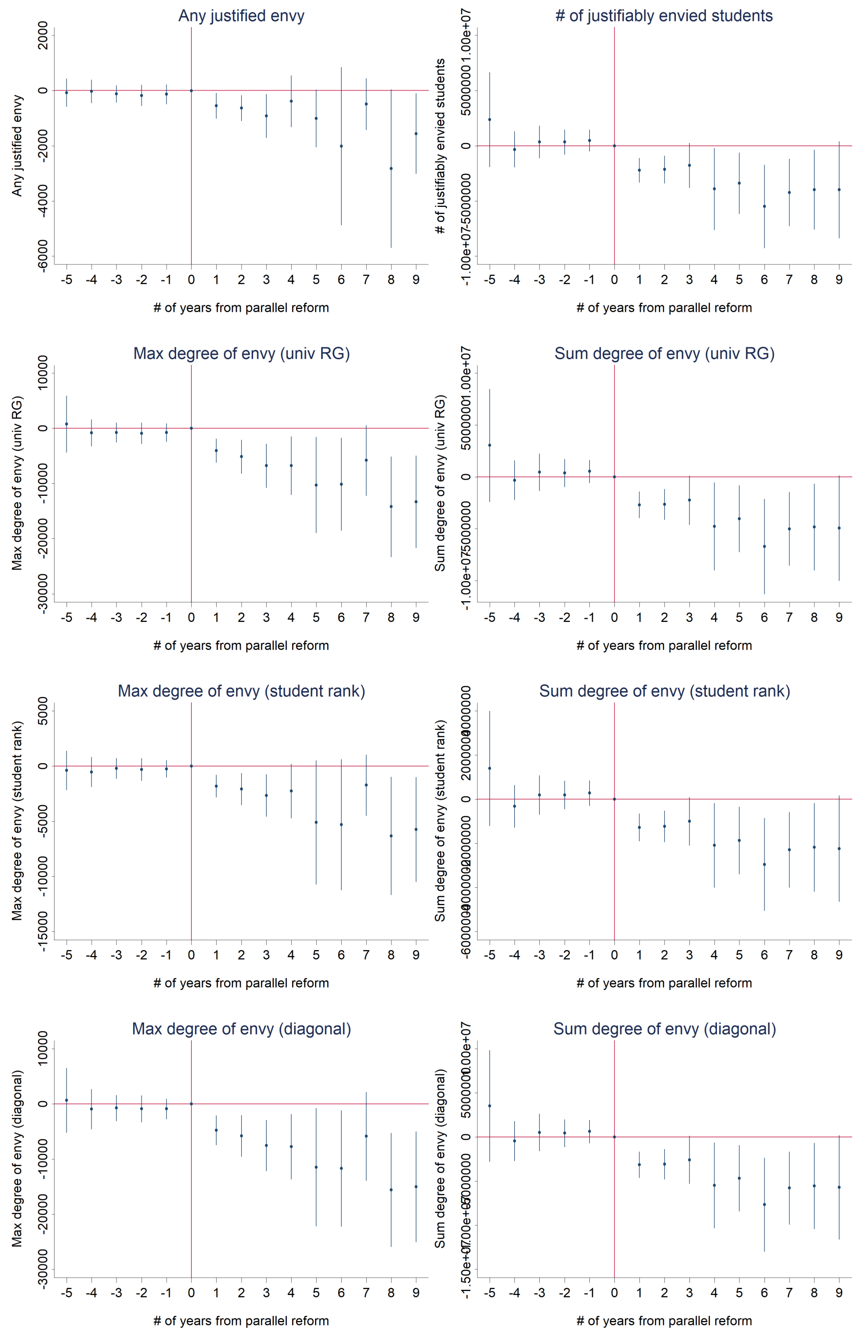
Table 12: Robustness Checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependant variable	Any justified envy (0/1)	Number of justifiably envied students	Max degree of envy (univ RG)	Sum degree of envy (univ RG)	Max degree of envy (rank)	Sum degree of envy (rank)	Max degree of envy (diagonal)	Sum degree of envy (diagonal)
Mean	0.806	357.2	1.467	402.9	0.870	168	1.742	455.9
SD	0.396	951.6	0.959	1126	0.740	576.7	1.032	1327
Panel A. Control for Timing Reform Dosage								
Parallel	-0.0419**	-208.3***	-0.408***	-257.0***	-0.178***	-121.2***	-0.484***	-298.8***
	(0.0207)	(25.59)	(0.0892)	(28.80)	(0.0448)	(14.77)	(0.105)	(33.68)
Panel B. Control for Partial Parallel								
Parallel	-0.0443*	-197.5***	-0.419***	-242.8***	-0.182***	-118.2***	-0.496***	-284.4***
	(0.0238)	(33.76)	(0.0906)	(38.47)	(0.0481)	(20.22)	(0.105)	(45.43)
Panel C. Exclude Provinces and Years with Tier 0 Pre-Admission Mechanism Reform								
Parallel	-0.0507**	-206.2***	-0.432***	-253.3***	-0.191***	-118.6***	-0.510***	-293.9***
	(0.0209)	(25.67)	(0.0902)	(27.71)	(0.0475)	(14.60)	(0.107)	(32.55)
Panel D. Drop highest and lowest majors from each university								
Parallel	-0.0815***	-176.4***	-0.438***	-208.4***	-0.277***	-106.8***	-0.512***	-247.1***
	(0.0235)	(29.85)	(0.0984)	(33.68)	(0.0640)	(16.84)	(0.118)	(39.02)
Mean	0.849	307.4	1.445	344.3	0.994	155.7	1.786	396.3
SD	0.358	821.4	0.864	978.1	0.805	531.5	0.964	1171

Notes: Each cell reports the coefficient of parallel reform from a separate OLS regression, including track fixed effects, province fixed effects, year fixed effects, reform timing controls, the total number of students participating in CEE in the track-province that year, and individual's score percentile. The mean and standard deviations for each outcome variable for the sample in Panel A, B, and C are presented at the top of this table. For Panel D, we present the new sample statistics below the coefficients. Robust standard errors clustered at the province level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

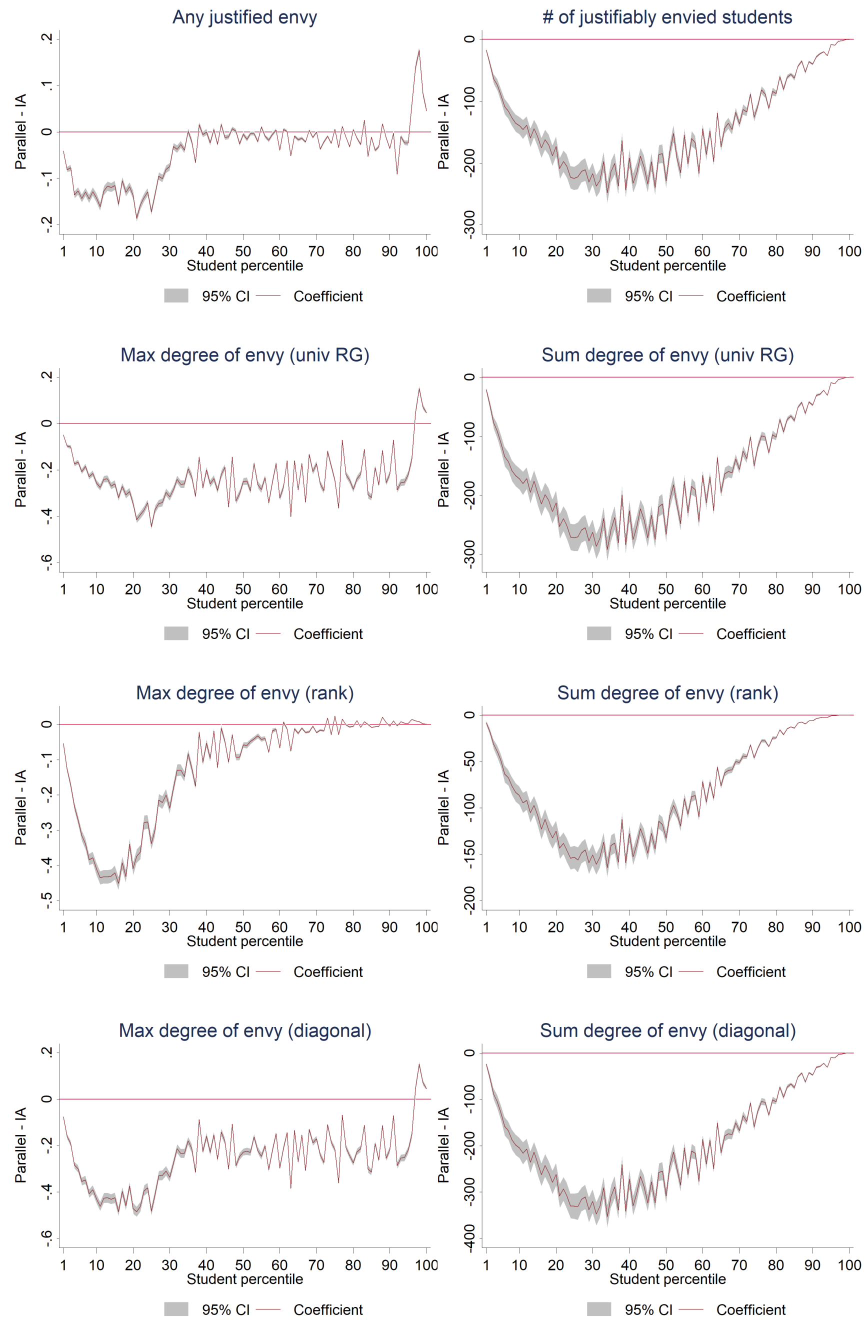
Appendix A: Supplementary Figures and Tables

Figure A1: Event Study Results on Normalized Envy Measures at the Matching Game Level



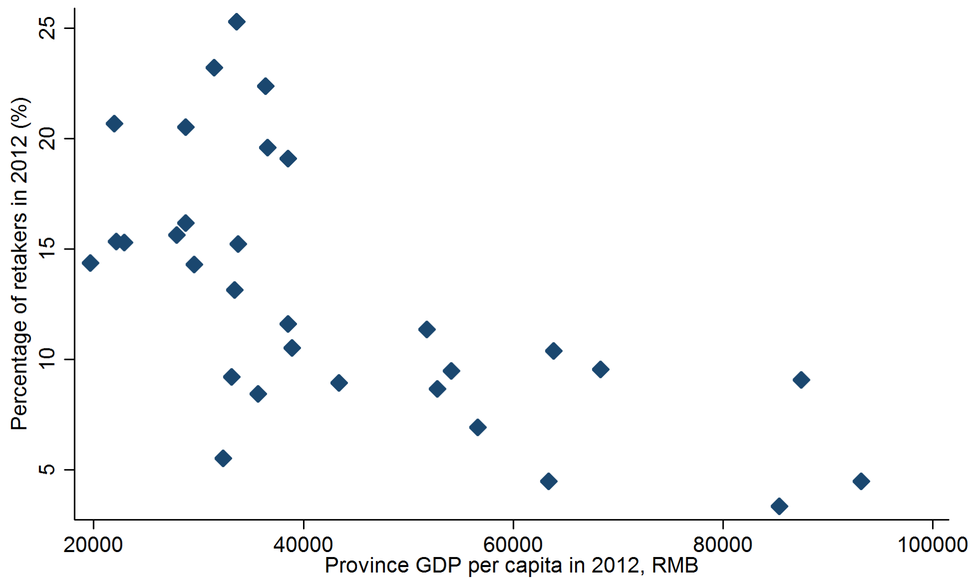
Notes: These figures present the event study coefficients of Equation 3. The outcomes from left to right, top to down are explained as follows. The first two outcomes are whether a student has any justified envy or not and the total number of students a student justifiably envies. The second row presents degree of envy in the dimension of university rank group. Maximum degree of envy (RG) equals 0 if no justified envy and equals own rank group admitted minus the best rank group for anyone who has a lower test score. Sum degree of envy (RG) equals 0 if no justified envy and equals the sum of degree of envy for all the envied students, i.e. adding up all the differences between envied student's rank group admitted and own rank group admitted. The third row presents degrees of justified envy in the student rankings dimension, which are similar to the previous two. The criteria used to assign a university into a rank group is the average admission score percentile rank that year. The graphs show that there are no pre-trends prior to the parallel reform.

Figure A2: Differences in Justified Envy Measures across Academic Performance Distribution



Notes: These figures present the differences in the average values of various justified envy measures along the CEE score distribution for matching games with IA versus parallel mechanisms. The solid line connects 100 coefficients, each from comparing the justified envy measure for students with CEE scores within a one percentile sample corresponding to the X-axis values. The grey areas show the 95% confidence interval. Note that in all our analyses, we limit our sample to those with a CEE score above the Tier I cutoff score. A lower percentile means a higher CEE score.

Figure A3: Scatter Plot of Provincial GDP per capita and Percentage Retakers in 2012



Source:

Table A1: Results from not assigning justified envy values to all individuals

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependant variable	Any justified envy (0/1)	Number of justifiably envied students	Max degree of envy (univ RG)	Sum degree of envy (univ RG)	Max degree of envy (rank)	Sum degree of envy (rank)	Max degree of envy (diagonal)	Sum degree of envy (diagonal)
Parallel	-0.0742***	-239.0***	-0.494***	-292.0***	-0.237***	-137.1***	-0.590***	-339.0***
	(0.0195)	(32.06)	(0.0867)	(34.93)	(0.0437)	(17.31)	(0.104)	(40.47)
Ex-interim	-0.0819**	215.4*	-0.399**	244.4	-0.203*	120.6	-0.417*	286.5
	(0.0319)	(128.1)	(0.157)	(146.7)	(0.107)	(75.29)	(0.225)	(173.3)
Ex-post	-0.0504	-119.8*	0.000516	-125.2	-0.0293	-70.49*	0.0650	-150.4*
	(0.0330)	(62.83)	(0.146)	(75.72)	(0.109)	(38.48)	(0.220)	(89.27)
Observations	3,373,755	3,373,755	3,373,755	3,373,755	3,373,755	3,373,755	3,373,755	3,373,755
R-squared	0.162	0.097	0.348	0.091	0.228	0.080	0.233	0.089
Mean	0.891	394.9	1.622	445.4	0.962	185.7	1.926	504
SD	0.312	993.1	0.875	1176	0.719	603.6	0.907	1386
effect/mean	-0.0833	-0.605	-0.305	-0.656	-0.247	-0.738	-0.306	-0.673
effect/sd	-0.238	-0.241	-0.565	-0.248	-0.330	-0.227	-0.650	-0.245

Notes: Each cell reports the coefficient of parallel reform from a separate OLS regression, including track fixed effects, province fixed effects, year fixed effects, reform timing controls, the total number of students participating in CEE in the track-province that year, and individual's score percentile. In this sample, we do not assign any justified envy values to those who would never justifiably envy anyone: those who were admitted to a top university in the highest rank group, and those who had the lowest test score in that matching game, i.e. the Tier I cutoff for that track-province-year. Robust standard errors clustered at the province level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A2: Compare Ningxia and National Effect Sizes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependant variable	Any justified envy (0/1)	Number of justifiably envied students	Max degree of envy (univ RG)	Sum degree of envy (univ RG)	Max degree of envy (rank)	Sum degree of envy (rank)	Max degree of envy (diagonal)	Sum degree of envy (diagonal)
				National				
Parallel	-0.0461**	-208.3***	-0.415***	-254.8***	-0.182***	-119.9***	-0.489***	-295.9***
	(0.0199)	(24.18)	(0.0848)	(26.74)	(0.0447)	(14.15)	(0.101)	(31.50)
effect/mean	-0.0572	-0.583	-0.283	-0.632	-0.210	-0.714	-0.281	-0.649
effect/sd	-0.116	-0.219	-0.432	-0.226	-0.247	-0.208	-0.474	-0.223
				Ningxia				
Parallel	0.00801	-33.44***	-0.0850***	-41.00***	0.0556***	-19.41***	0.0198	-47.44***
	(0.00853)	(3.118)	(0.0152)	(3.607)	(0.0132)	(1.876)	(0.0175)	(4.259)
effect/mean	0.0110	-0.477	-0.0821	-0.547	0.0860	-0.629	0.0160	-0.561
effect/sd	0.0180	-0.218	-0.110	-0.234	0.0860	-0.201	0.0227	-0.226

Notes: Each cell reports the coefficient of parallel reform from a separate OLS regression, including track fixed effects, province fixed effects, year fixed effects, and reform timing controls. Robust standard errors clustered at the province level in parentheses. *** p<0.01, ** p<0.05, * p<0.1