Personalized Advising for College Match: Experimental Evidence on the Use of Human Expertise and Machine Learning to Improve College Choice

Xiaoyang Ye
Brown University

Abstract

This paper studies how machine learning predictions can improve the effectiveness and efficiency of college-choice advising. When applying to colleges, students often incorrectly predict their admission probabilities. Personalized advising can effectively reduce prediction mistakes but is not scalable. Using a large-scale field experiment under centralized admissions, I showed that personalized advising substantially improved college access and match without changing students' college preferences. Machine learning predictions reduced the human labor of the intensive data analysis needed for advising but achieved similar treatment effects to conventional expert advising. A supplemental survey experiment decoded how human expertise and machine learning improved college-choice decisions.

Keywords: College choice, academic match, behavioral intervention, machine learning

JEL Codes: I22, I23, I24, C93, D91

Data Availability Statement: The first part of this paper uses confidential administrative data that other researchers could apply directly to the Ningxia Department of Education. The author commits to providing all the codes, non-confidential data (including the experimental data of the second part), and guidance as to how to apply for the confidential data, as well as ensures that the results are fully replicable.

Disclosure Statement: This research was made possible with generous funding from the National Academy of Education/Spencer Dissertation Fellowship, China NSF (71613013), China Ministry of Education (16JJD880001), China Dexinren Grant on Education and Poverty, the University of Michigan, and Princeton University Data Driven Social Science Initiative. The research was approved by the University of Michigan IRB (HUM00127559).

1 Introduction

The decision of whether and where to attend college is widely recognized as a critical moment that can have a profound impact on students' lives. Unfortunately, many students make suboptimal college choices during the complex transition from high school to college, which results in them attending an academically undermatched college (Bowen, Chingos and McPherson, 2009; Hoxby and Avery, 2013; Smith, Pender and Howell, 2013; Dillon and Smith, 2017; Black, Cortes and Lincove, 2020). College undermatch has large, negative impacts on students' college and labor market outcomes (Howell and Pender, 2016; Ovink et al., 2018; Dillon and Smith, 2020), offsetting other policy efforts to improve college-going outcomes through academic preparation and financial aid (Page and Scott-Clayton, 2016). Over the past decade, research and policy interventions have emerged focused on improving students' college choice decisions – ranging from low-touch information interventions to intensive personalized advising programs.[1]

One of the decisions that result in academic undermatch is to apply to college based on inaccurate predictions of the probabilities of college admissions due to lack of information and/or guidance (Hoxby and Avery, 2013; Kapor, Neilson and Zimmerman, 2020; Arteaga et al., 2021; Mulhern, 2021). Personalized advising can effectively reduce prediction mistakes and improve college match; however, those programs are difficult to be implemented at scale.[2] As a common challenge to social policy programs, the scale-up problem results from the same underlying constraint: Some "inputs" to a program are in limited supply (Davis et al., 2017; Muralidharan and Niehaus, 2017). Advisers – the key input to college-going advising programs – are inelastically supplied in quantity and quality, and often come with high costs. Since effective college-going advising requires intensive guidance, instruction, and assistance and we may not be able to hire and train enough qualified advisers to

[1] Recent summaries include Thaler and Sunstein (2008); White House (2014); Castleman, Schwartz and Baum (2015); Lavecchia, Liu and Oreopoulos (2016); Page and Scott-Clayton (2016); Castleman (2017); French and Oreopoulos (2017); Damgaard and Nielsen (2018); J-PAL (2018); and Schmidt and Park (2021).

[2] Advising examples include Bettinger et al. (2012); Carrell and Sacerdote (2017); Oreopoulos, Brown and Lavecchia (2017); Oreopoulos and Petronijevic (2018); Bettinger and Evans (2019). In contrast, light-touch information and nudge interventions can be provided to a large number of students with a low cost but may not be sufficiently effective in reducing college undermatch (Hoxby and Turner, 2013; Bergman, Denning and Manoli, 2019; Evans and Boatman, 2019; Hyman, 2019; Avery et al., 2020; Gurantz et al., 2020*a*; Bird et al., 2021).

provide such advising (Bettinger and Evans, 2019; Gurantz et al., 2020*b*), personalized advising is unlikely to be scalable as a system-level policy intervention to serve a large number of students.

In this paper, I study a new policy solution that uses machine learning predictions to scale up an effective personalized college application advising program to improve student-college match and reduce inequality in college access. In prior work, I developed a personalized advising program to help students make informed college choice decisions (Ye, 2020). Personalized advising provided students with the "human expertise" to make accurate predictions of college admissions probabilities and to submit college applications based on the predictions. However, this effective advising program had the limitation of requiring a a large supply of expert advisers. This paper proposes and examines an effort to scale up the personalized advising through increasing the intensive margin of labor supply by using big data and machine learning algorithms to simplify the prediction of college admissions probabilities – the most complex and time-consuming step during the personalized college-choice advising.

To estimate the causal effects of machine learning-assisted personalized advising on college choice behaviors and admissions outcomes, I conducted a large-scale randomized controlled trial (RCT) among the universe of high school graduates in 2017 in Ningxia, one of the poorest provinces in China. China is an ideal setting to study behavioral interventions in college choice as it has the largest centralized college admissions market in the world: Each year, about ten million Chinese high school graduates apply to colleges; and more than three million students undermatch because of their college choices. As the centralized system eliminates behavioral barriers for information and simplifies application process, this paper is able to identify students' strategic decisions in college choice. Furthermore, I am able to credibly identify the impacts of college choice behaviors on admissions outcomes, because the variation in admissions outcomes is solely determined by students' college choice behaviors when holding their college entrance exam scores constant.[3]

In the second year of the *Bright Future of China Project-Ningxia* following Ye (2020), with

[3]In contrast, decentralized admissions consider both academic achievement and other confounding factors including extracurricular activities, athletic abilities, and personal qualities, some of which are not observable to researchers and thus may result in omitted variable bias.

the close collaboration with Ningxia Department of Education, I used a student-level stratified randomization design. As summarized in Figure 1, I randomly assigned students into one of the three groups: (1) 5,647 students were provided access to the machine learning-assisted personalized advising; (2) 5,370 students were provided access to a low-touch "business as usual" advising group; and (3) the remaining 43,038 students served as the control group.[4] In the "machine learning" group, expert advisers used the assistance of machine learning predictions and relevant data analytics to provide students with conventional personalized advising. The "business as usual" advising provided a generic college application guide.

Results indicate that the personalized advising assisted by machine learning substantially improved college access and match, which closely mirrored the effect of the pure expert advising in the previous year but dramatically increased the advising productivity. Compliers of the machine learning-assisted advising were admitted to colleges where entrance exam score quality measure is 0.6 standard deviations higher.[5] Heterogeneity analyses using both conventional linear regression and Causal Forests suggest that students from economically disadvantaged families benefited more than their advantaged peers from that personalized advising.

By analyzing college choice behaviors using the unique data on students' college applications, I found that machine learning advising nudged the treated students to be more likely to correctly predict college admissions probabilities and act on these predictions. As the advising program only focused on improving students' abilities of making accurate predictions of college admissions, students in the machine learning intervention group increased college access and match without changing their preferences for college attributes and major choices.

To further understand how human expertise machine learning algorithms improved college-choice decisions, I decoded the "machine learning black box" using an incentivized survey experiment. Among a nationwide sample of 2,542 Chinese high school graduates in 2020, participants

[4] The number of students in the two treatment groups were determined by the estimated individualized advising capacity and the expected take-up rate of 20% in each group.

[5] With a sample of 39,772 students in three application periods, Arteaga et al. (2021) find a treatment-on-the-treated effect of 0.1 s.d. on test score value added from providing students with live feedback on assignment probabilities in centralize college admissions.

were asked to submit college applications for the same hypothetical applicant, whose admissions outcomes was determined using the actual college admissions results one month later. I randomly provided students with different access to various machine learning elements, including human expertise in college choice, data, and machine learning prediction information.

Results were consistent with those in the real-world field experiment. Similar to the null effect of the "business as usual" advising, access to the human expertise regarding college choice strategies did not improve college choice behaviors and outcomes. However, the combination of human expertise and data largely improved college choice quality and admissions results, which demonstrated that the estimated positive effects of machine learning-assisted personalized advising on college access and match was not a coincidence. Additionally, the result that the pre-registered "AI reference" application portfolio beat all the 2,542 participants suggested that machine learning has the potential to support human decisions in complex optimization problems.

This paper makes several contributions to the literature and policymaking. First, this paper provides new, compelling evidence of that a behaviorally-designed, intensive intervention using a combination of customized information and personalized assistance substantially improves students' college choice behaviors (particularly, their use of prediction strategies) and thus their college access and match outcomes. The intervention design in the *Bright Future of China Project-Ningxia* builds on and expands many prominent approaches in the literature, particularly personalized advising/counseling (Bettinger et al., 2012; Castleman, Owen and Page, 2015; Carruthers and Fox, 2016; Carrell and Sacerdote, 2017; Oreopoulos, Brown and Lavecchia, 2017; Page et al., 2017; Castleman and Goodman, 2018; Evans and Boatman, 2019).[6] While recent interventions aiming to influence students' college lists were ineffective (Phillips and Reber, 2019; Fesler, 2020), this paper shows the importance of providing structural, data-based guidance in helping students make accurate predictions and improved college choices. Moreover, I also show that taking sufficient time to form a thoughtful college choice plan improves students' college access and match. This finding

[6] I also considered other light-touch informational interventions, but most of them were not suitable in the centralized admissions context. Bird et al. (2021) show that these light-touch interventions, though successful in small scale projects, may not be effective at scale in the U.S. higher education context.

suggests that nudging students to "think slowly" may have desirable behavioral consequences (Kahneman, 2011; Heller et al., 2017).

Second, this paper fills the literature gap of limited evidence on college-going interventions from centralized admissions systems, or broadly K-12 and higher education centralized admissions systems. Existing literature concentrates on the higher education markets in the U.S. and Canada; see summaries in White House (2014); Page and Scott-Clayton (2016); French and Oreopoulos (2017); J-PAL (2018). We know relatively little about how personalized advising works in centralized systems (Dinkelman and Martínez A, 2014; Hastings, Neilson and Zimmerman, 2018; Peter, Spiess and Zambre, 2018). This paper provides new evidence on the impact of college application assistance from the largest centralized college admissions market in the world. Centralized admission is widespread across countries in both K-12 and higher education (Neilson, 2019).[7] Centralized admission systems streamline and simply the application process; however, they require strategic decision-making that creates a barrier for many students. A recent study by Arteaga et al. (2021) shows that providing application feedback information improves school and college choices. The effective intervention described in this paper supports students, especially disadvantaged students, in the search and selection of colleges to apply to.

Third, this paper casts important implications on the scale-up of social policy programs. Personalized advising is effective in improving college access and match, but they are not easily scalable. Existing intensive college counseling studies have only covered a small number of students. For example, Carrell and Sacerdote (2017) provide a college coaching/mentoring program at the cost of $300 per student, but the program has only 871 treated students in six high school graduation cohorts. Oreopoulos, Brown and Lavecchia (2017) evaluate the Pathways to Education Program in

[7]In countries like Brazil, Chile, China, Germany, Greece, India, South Korea, Turkey, and the United Kingdom, college admissions operate through national exams and a centralized application and admission system. Many American colleges are starting to use the Common Application, a platform through which students may submit the same application to as many colleges and universities as they like. Many K-12 school admissions are centralized, such as in Amsterdam, Boston, Paris, and New York (Hafalir et al., 2018), as well as in all Chinese cities. In a recent discussion, Goodman and Rucinski (2018) propose a centralized testing and admission policy for Boston's exam schools that assigns students based on universally-taken test scores would largely increase the number of Black and Hispanic students in the exam schools. In the centralized K-12 school choice in New Haven, CT, Kapor, Neilson and Zimmerman (2020) find that a large proportion of families make incorrect predictions of school assignment probabilities.

Canada that has served only 1,274 students in 8 years. This paper proposes and tests a new intensive margin solution to increasing the labor supply of college choice "human expertise" services, which may be applied in addressing many other policy problems. Bird et al. (2021) provided FAFSA nudge campaigns, including college advising, various information contents and delivery method, and numerous reminders, at scale to 800,000 students, but found no effects. One possibly is that the generic, less personalized nudges are difficult to change student behaviors. Scaling up more intensive, individualized interventions would matter for student behaviors and outcomes.

Fourth, this paper contributes to the recent literature on applying machine learning to prediction policy problems (Kleinberg et al., 2015; Mullainathan and Spiess, 2017). When past data are available to learn from, the link between predictions and decisions is clear, and expertise in understanding the decision process is applied, machine learning algorithms show considerable potential for improving predictions and productivity (Chalfin et al., 2016; Kleinberg et al., 2017).[8] Data-based decision-making has been rapidly growing in both K-12 and higher education, such as computer/technology-assisted instruction (Muralidharan, Singh and Ganimian, 2018; Taylor, 2018), digital tutoring (Burch, Good and Heinrich, 2016), learning analytics (Daniel, 2015), and predicting college application, enrollment, and success (González and DesJardins, 2002; Herzog, 2006; Acharya and Sinha, 2014; Aulck et al., 2016). In a study of technology-based college coaching using online exercises and text and email messaging, Oreopoulos and Petronijevic (2018) find no effects of the technology-based intervention and conclude that "future technology-based interventions should aim to provide proactive, personalized, and regular support." Machine learning or data-based prediction has the potential to offer personalized assistance, perhaps better than human experts. Lechner and Smith (2007) examine the efficacy of caseworkers in allocating individuals to government programs, and find that statistical treatment rules do substantially better. Burkhardt et al. (2018) find that medical school enrollment predictions using the enrollment management model

[8] Machine learning is not a master key to all policy prediction problems without human expertise in understanding the problems. For example, McKenzie and Sansone (2017) study the prediction of outcomes for entrants in a business plan competition in Nigeria and find that machine learning methods do not offer noticeable improvements. The main reason is that the overall predictive power of both human judges and prediction models is very low, which means the key variables of the decision-making are not being measured. Bird et al. (2020) observe similarly high levels of model accuracy between the simplest and most complex models in predicting community college student outcomes.

were at least as accurate as the expert human estimates, and in specific populations of interest more accurate. Mulhern (2021) finds that personalized admissions information largely shifts students' college choices. This paper adds experimental evidence on the effectiveness of machine learning for improving prediction and advising in the college-going behavioral interventions.

2 Background

2.1 Why Do Centralized Admissions Need (More) Accurate Predictions?

College choice, whether and where to go to college, is one of the highest stakes decisions in life. Each year, millions of high school graduates all around the world make their college choices through either a decentralized system or a centralized system with national college entrance exams (Neilson, 2019). Recent college-going intervention literature focuses on decentralized college admissions systems and suggests that relatively inexpensive information provision and application process simplification can substantially improve college access and match. These two policy levers have already been institutionalized in the centralized systems.[9] The centralized system largely simplifies the application process: after taking the national entrance exam, students only need to submit a rank-order application list of colleges and majors.

However, students in centralized systems still face behavioral barriers in constructing their rank-order application portfolios. In particular, as noted by Lavecchia, Liu and Oreopoulos (2016), students may make mistakes with too little information or with too many options. The college choice model assumes that students make their optimal choices by comparing the benefit-cost tradeoffs between college and major options (e.g., Manski and Wise, 1983; Kane, 1999; Long, 2004; Perna, 2006; Jacob, McCall and Stange, 2018). One the one hand, lack of information, misinformation, or

[9] The centralized admission mechanism, considered to improve efficiency, welfare and match (Gale and Shapley, 1962, Balinski and Sönmez, 1999; Abdulkadiroğlu and Sönmez, 2003), has long been adopted in many markets, including college admissions in many countries and in some U.S. K-12 school choices (Abdulkadiroğlu, Pathak and Roth, 2005; Pathak and Sönmez, 2013; Machado and Szerman, 2017). For example, while there is a growing literature documenting that mandatory entrance exam and automatic score sending in decentralized systems (particularly in the U.S.) improve college access and enrollment (Klasik, 2013; Bulman, 2015; Hurwitz et al., 2015; Pallais, 2015; Goodman, 2016; Hyman, 2017; Hurwitz et al., 2017), centralized admissions systems eliminate the barriers in testing for students.

unawareness about college features – cost, return, curriculum, major, and special programs – and admissions policies makes it impossible for students to correctly compare the tradeoffs. Smith, Pender and Howell (2013) and Dillon and Smith (2017) suggest that students who have better access to information on college options and the college going process (e.g., from parents, networks, and schools) are less likely to undermatch.

On the other hand, students may also have the overchoice problem when facing too many options. Students may have limited cognitive capacity and attention when evaluating a large number of choices and identifying the best fit options, for example, identifying a short list of reach, peer, and safety colleges from thousands of colleges with multidimensional information. Kapor, Neilson and Zimmerman (2020) and Arteaga et al. (2021) show that applicants often have biased beliefs about the admissions probabilities when applying to K-12 schools or colleges in centralized admissions systems. Mistakes in the predictions result in large risks of non-placement or being admitted to less desirable options.

A wise college choice strategy is a function of students' test scores, their preferences and valuations for each college, their predicted admission probability, and other individual idiosyncratic factors. In centralized admissions, the admission result is solely based on a student's rank-order list and entrance exam score. It is then critical for students to thoughtfully select a smaller number of colleges from the available options to apply to and rank them in an order that maximizes their expected outcomes. Similar to the expert's advice in decentralized systems as described by Hoxby and Avery (2013), applying to a mix of reach, peer, and safety colleges and ranking them in an ascending order of predicted admissions probabilities is also an appropriate strategy to maximizing the chances of being admitted to higher-quality colleges while minimizing the risks of non-placement.[10] To construct such a portfolio, students need to make an accurate prediction of the admission probability for every considered college, which requires an understanding of the admissions mechanism, decision-making skills, and the use of historical admissions data for

[10] There are large income and racial gaps in this college application strategy. In the U.S., Hoxby and Avery (2013) find that, while high-income students generally follow expert's advice to apply to a mix of reach, peer, and safety colleges, the vast majority of low-income high achievers do not apply to any selective colleges. Loyalka, Wu and Ye (2017) and Campbell et al. (2021) find similar results in China and the U.K.

predictive analytics.

Several common institutional features and barriers in centralized admissions systems emphasize the role of accurate prediction and strategic application. First, admissions are often implemented using an application list with a restricted length in many real-world examples (Arslan, 2018; Chen and Kesten, 2017). For example, students in Ningxia (the study sample of the first field experiment in this paper) could only apply to four selective colleges and six majors within each college.[11] Under the Deferred Acceptance mechanism, but with restrictions on the application, truthful revelation of preferences is no longer a good strategy. Second, many centralized matching systems only match subjects (e.g., students) to *at most* one option, which imposes a high risk of all applications being rejected when students aim too high. Students have to evaluate the admission probability for each college based on admissions outcomes (cutoff or median scores) in prior years. Third, students must rank their applications in order. Without accurate information about college quality (or other individual preferences), and without precise predictions of admission probabilities, students may make mistakes in ordering, e.g., listing a safety college at the first choice, or listing a reach college at the last choice (Kapor, Neilson and Zimmerman, 2020). Lastly, centralized admissions may operate in a very short period of time.[12] In order to conduct a thorough search and assess college fit, students have to very efficiently search for and analyze a large volume of college/major information (thousands of colleges) from sources.

2.2 Context

Chinese college admissions. This paper studies college applications and admissions in the Chinese province-level centralized college admissions system. At the end of high school, students take the annual College Entrance Exam (CEE). Students then compete with peer applicants within the same province and STEM/non-STEM track for college-major spots that are predetermined by

[11]Even in decentralized systems where students could potentially apply to as many colleges as they want, costly applications (e.g., a complex process and application fees that limit the number of applications) require students to apply in a sophisticated way when choosing their final applications.

[12]Most Chinese students only start to seriously think about college choice after they know their CEE scores. They only have three to five days to submit college applications.

each college across the country.[13] Students submit their college application lists (typically 4-10 colleges in each institutional tier, 4-6 majors for each college, varying by provinces and institutional tiers) in the Department of Education online system.[14] Admission is solely determined by students' CEE scores and their college-major applications. Each student is matched with *at most* one college-major through a parallel mechanism (like the Deferred Acceptance mechanism, as discussed in Chen and Kesten, 2017). Students who decline the admission offer or are not admitted to any college have the options to retake the CEE in the next year, move onto the labor market, or seek to study abroad.

As discussed in the previous subsection, Chinese students also need to strategize on the basis of predicted college admissions probabilities. Similar to the expert advice as described by Hoxby and Avery (2013), they could apply to a set of reach, match, and safety colleges to maximize their admissions opportunities by a reach or match college, and to minimize their chances of being rejected by all of the colleges to which they have applied. Since college admission is uncertain and risky - with a limited number of choices and each student admitted to at most one college - students must "game" the college application strategically with accurate predictions. Additionally, students have to choose college and major simultaneously. The match process is college-then-major, which perplexes the college choice decision-making process, even though the application process itself is simple and most information is available to all the students.

Setting and data. Through the research-policy partnerships with the Ningxia Department of Education and the Ningxia Education Examination Board, the provincial centralized administration office of the College Entrance Exam and college admissions, I collected the student-level administrative data for the universe of the 2017 high school graduation cohort in Ningxia province. Figure A.1 shows the geographic location of Ningxia - one of the lowest income provinces located in

[13]Students choose one of the two tracks one or two years before taking CEE. They take four subjects: Mathematics, Chinese, English, and track composite. The STEM track composite includes physics, chemistry, and biology. The non-STEM composite includes history, social studies, and geography.

[14]College application and admission proceeds by institutional tiers and students' eligibility for applying to different tiers is determined by their CEE ranking percentile, e.g., Tier 1 includes the nation's elite colleges and only the top 10% to 20% students can apply to. All of the relevant information, including college admissions results in prior years, and information about tuition, location, and quota are publicly provided to students by the Department of Education. I show and explain a typical Chinese college application form in Subsection C.1.

northwestern China. Using data of the entire population of applicants in a college-student matching market enables me to identify students' college application behaviors (strategies and preferences), admissions outcomes, and enrollment decisions.

The confidential student-level data include student demographics and high school attendance records, CEE scores, full rank-order applications data, and admissions results. I discovered and cleaned a new dataset of the college application submission time, which I used to analyze the intervention's effects on the timing of applications. I also constructed detailed student-college-major-choice level data (over 1 billion observations) to develop the machine learning algorithms that predict the admission probability of every college-major option in each possible application rank order for each student.

I merged the college-major level information, consisting of address, tuition, quota, prior-year admissions scores, with the student-level data in order to study their college choice strategies and preferences. The college-major level information was the same as that provided to students during the college application period by the Ningxia Department of Education in printed books.[15] I used student level and college-major level data to create a family of measures of college access and match outcomes, as well as college choice behaviors. I also obtained additionally college-major data (e.g., national college rankings and track-tier admissions cutoff scores) from external sources.

2.3 Motivating Evidence on the Importance of College Choice Strategies

Following the undermatch literature (e.g., Hoxby and Avery, 2013; Smith, Pender and Howell, 2013; Dillon and Smith, 2017) and my previous work (Ye, 2020), I constructed an index of student-college academic match using principal-component analysis of five college quality measures, including college admissions scores in Ningxia in 2017 (median, mean, minimum) and national college quality measures (standardized score and ranking percentile).[16] In centralized admissions,

[15]The necessary information is available to all students. But the delivery using printed books imposes high search and analytical costs for students to make optimal choices and decisions.

[16]Using college admissions data from 1996-2017 and administrative data on institutional resources for every college in China, we build a national college ranking of all Chinese colleges, which is now published at siminedu.com to assist all Chinese high school graduates in their college choices.

the between-group difference in college undermatch is solely due to college choice behaviors when controlling for CEE scores.

To illustrate the importance of college choice strategies, I first identify the poverty gap in college match using the following linear model:

$$Y_i = \beta_0 + \beta_1 * Rural_i + \gamma * X_i + \varepsilon_i \tag{1}$$

where β_1 measures the rural-urban gap in the college match outcomes Y_i, holding individual covariates X_i (CEE score, demographics) equal. Given the limitation of the administrative data, I used rural *hukou* (household registration, the primary source of income inequality in China) as a proxy for poverty.[17]

Low-income students from rural families are much more likely to undermatch compared with their higher income peers. Using the sample of students who submitted their college applications and were not in the treatment groups in 2017, Column 3 of Table 1 shows that rural students are admitted to a college with 0.107 standard deviations ($p<0.001$) lower quality than urban students with the same CEE scores and demographic characteristics.[18] Since students in the same high school may share the information and support from the school, controlling for high school fixed effects reduces more than one-third of the gap (β_1=-0.063 in column 7). However, the sizable rural-urban gap in college match persists and is attributable to students' different college choice behaviors.[19]

I next predict admissions outcomes based on college choice behaviors. I constructed a series measures of strategies and preferences using students' full applications data. Appendix Subsection C.1 provides a detailed description of these measures and underlying behavioral rationales. I add the six principal-component indices (standardized) stepwise to Model (1). Data-based targeting

[17] Note that Ningxia is one of the poorest provinces in China but there still exists large poverty gaps within Ningxia. Results are qualitatively identical when using other measures of family income or disadvantaged background, e.g., female or minority.

[18] Table B.1 shows that, including those who did not apply to any college (and were assigned the track-tier lowest admission score), the raw rural-urban gap in the college match index is -0.134 standard deviations. Rural students end up with colleges on average 0.149 s.d. lower quality when controlling for CEE score and demographics. Female and minority students are also more likely to undermatch, while repeaters are better matched than first-time exam takers.

[19] All the high schools, located in urban districts, have a mix body of students from rural or urban families. Controlling for class fixed effects or neighborhood fixed effects does not change the results once we control for school effects.

strategies are the core of college application expertise, which enables students to make accurate predictions of admissions chances and to apply for a targeted set of reach, match, and safety colleges, as well as rank them in an appropriate order in the application list.

The first two columns of Table 1 show the sample means of each college choice behavior index. Rural students are less likely than urban students to use the targeting strategy and follow the general advice (e.g., fill in all the major applications within each college). Moreover, rural students prefer colleges with lower tuitions and larger admissions quotas, and in-province colleges that would limit other high-quality college opportunities (Hillman, 2016; Ovink et al., 2018).

Column (4) shows that a one standard deviation increase in the use of the *prediction-based targeting strategies* is statistically significantly correlated with a 0.2 s.d. increase in the quality of the admitted college. The correlation is stable when controlling for other college choice measures in column (6). The other measures are also correlated with admissions outcomes, but in much smaller magnitudes. Comparing the changes in the rural-urban gap (β_1), I find clear evidence that prediction-based targeting strategies explain the largest proportion of the variation in the outcome (40 percent), controlling for CEE score and demographics. This result remains unchanged when I control for school fixed effects in column (10). Targeting strategies explain about half of the rural-urban gap in college match.[20]

Overall, as displayed in columns (6) and (10), if rural students have the same college choice behaviors as urban students, the rural-urban gap reduces by more than 60 percent.[21] The use of targeting strategies requires human expertise in understanding admissions mechanisms as well as sophisticated data analytics. The key component of these targeting strategies is to accurately predict college admissions probabilities. I now turn to the randomized experimental design to examine the effectiveness of using machine learning predictions to improve the efficiency of expert personalized advising for college match.

[20]Results from Oaxaca-Blinder decompositions show that targeting strategies explain more than 80% of the rural-urban gap in college match that is explained by all the six college choice behavior measures.

[21]We should note that the six behavior measures included in the analyses do not fully capture a student's college choice strategies and preferences. Furthermore, the "optimal" application plan is not based on a single indicator, but a compound of a variety of strategies and preferences. Identifying an optimal college application plan is still an open question for future research.

3 Experimental design

3.1 Using Machine Learning to Improve Personalized Advising Efficiency

As part of the *Bright Future of China Project*, I designed a structured guide to help students who lack information or guidance to navigate the college application process. I conducted a pilot experiment in 2016 that provided proof-of-concept evidence that information and guidance about predicting admissions probabilities improve students' college choice behaviors and admissions outcomes (Ye, 2020). In particular, personalized advising provided by expert advisers effectively guided students make accurate predictions of admissions chances using a sequence of data analyses. However, performing such data analyses is labor-intensive and time-consuming, which limits the potential to scale up the personalized advising program.

I propose to use machine learning predictions to help the advisers provide personalized advising more efficiently and effectively. Given the solid evidence that an effective college choice advising relates to making accurate predictions, machine learning algorithms can reduce the time that an expert adviser needs to serve each individual student by providing automatic data analytics and predictions results. Machine learning algorithms mirror the predictions that a skillful expert or student would make in the college choice process; that is, learning from a history of past admissions outcomes to predict the admissions outcome at each college.

Predicting college admissions is core of the personalized advising program. An expert or a student's subjective prediction is often imprecise, yet still useful when qualitatively correct. For example, we may use a single criterion such as the five-percentile bandwidth to define reach, match and safety colleges (Hoxby and Avery, 2013). Using machine learning algorithms, one can predict the admissions probability of each college using multi-dimensional factors such as different measures of admissions scores, quotas, and college/major features. Even two safety colleges can differ substantially in their actual *ex ante* admission probabilities. When machine learning excels at precisely predicting *objective* admissions probabilities more than the human instruction approach, why not use it for better decisions?

Training the prediction model. Machine learning was used as a "black box" methodology that provides the prediction outcomes for the personalized advising program. I used the unusual complete data on student applications in a whole college matching market, thanks to our long-standing research-practice partnership with the Ningxia Department of Education. I then applied machine learning algorithms to generate precise predictions for each student for each college: notably, each type of students with the same CEE score, gender, and race (gender and race may affect eligibility for special programs or majors). Because of the detailed application data, I can even predict the admissions probability for each student at each college-major option of every rank order in the application list.[22]

When a skillful expert or student makes her "human predictions," it is difficult to consider many factors simultaneously, such as minimum/maximum/median/mean admissions scores of a college in the past few years. All of these "human predictions" are subjective approximations (e.g., reach vs. safety, or unlikely vs. likely). This problem is much simplified in a machine learning algorithm. Similar to Kleinberg et al. (2017), I provided "the machine" with a set of student-college-major-order level data from the 2016 cohort, each observation consisted of a set of input features, including admissions scores and corresponding ranking, quota, order, student's CEE score, as well as an outcome (admission=1) to be predicted. I then chose a prediction model (algorithm) to minimize the loss function that generates accurate out-of-sample predictions. I used Random Forest, one of the most commonly used supervised learning algorithms, which builds on a collection of decision trees to generate the predictions.[23] Figure A.2 shows feature importance for predictions. As expected, a student's ranking within the province-track (*paiming*, like an equated CEE score), her normalized CEE score (*normalized_zongfen*), and the rank order of the applications (*zyno*) carry the largest importance shares. A student's subject scores and a college's admissions

[22]In other cases, when we do not have such detailed data, using admissions scores, which are publicly available to students when they apply to college, still generates relatively precise predictions. Without a more sophisticated algorithm to help students evaluate the expected returns and risks of placing a college-major at different rank orders, the small difference of listing a college-major at different orders does not have large impacts on the predicted admissions.

[23]Assessing the performance of different machine learning algorithms vs. human instruction on the same set of students and applications is out of the scope of this paper. Subsection C.3 provides descriptions of model comparisons from a companion project.

scores in the previous year are also correlated with the predicted admissions, while other features are not that important. To avoid over-fitting, I randomly partitioned the data into an 80% training set and a 20% test set. Cross-validation shows that the prediction accuracy was 94.3% (95% confidence interval: 94%, 94.5%; suggesting a high accuracy) with a sensitivity of 83.4% and a specificity of 98%.[24] Since the goal of prediction is college admission, the model is conservative in predicting the admissions among students' reach colleges.

3.2 The Intervention: Machine Learning-Assisted Personalized Advising

The machine learning-assisted advising. In the field, the intervention provided to treated students was machine learning-assisted online personalized advising, which followed the pure human advising program in 2016. Expert advisers guided students in three main procedures: (1) providing a comprehensive and reliable guide to help them search for college and major information; (2) using data analysis to compare a short list of colleges and majors by identifying college types based on predicted admissions probabilities; and (3) instructing the strategies to make optimal college choice decisions. The ultimate goal was to help students make accurate predictions of admissions chances and to select several targeted colleges out of more than 3,000 colleges. Subsection C.2 provides detailed descriptions of the personalized advising.

The personalized advising program was implemented using WeChat, the most popular message App in China. Two expert advisers were assisted with machine learning predictions and two additional data reports. First, the adviser used a program to automatically equate CEE scores for students, which was done manually in 2016. The adviser also provided treated students a short list of reach, match, and safety colleges, which aimed to reduce a student's search cost. Figure 2 compares the manual data analysis in the traditional advising with the automatic data report. This step reduced advising time from hours to seconds.

Next, the adviser asked treated students to provide a rank ordered list of colleges and majors

[24] Sensitivity and specificity (and accuracy) are common characteristics of the model prediction performance. Sensitivity is the proportion of observed college admissions that were predicted to be admitted by the model, and specificity is the proportion of observed college rejections that were predicted to be rejected.

that they considered applying to. The adviser then returned the predicted probabilities of each college-major in the candidate list. Figure 3 present screen shots of the interface that the adviser used during the personalized advising. Figure A.3 shows the conversations between the adviser and the student through the online message App. Students picked their final application lists consisting of a group of reach, peer, and safety colleges based on individual preferences and completed their applications.

This intervention design combined human expertise and algorithmic judgment in order to minimize potential errors and to make the best possible college application. During the advising, the main task of the experts was to help students check and fine-tune their candidate list. The majority of the conventional advising in 2016 was then replaced by machine learning (and Stata) prediction results in 2017. The advising productivity was greatly improved. Furthermore, as students might have heterogeneous preferences for college attributes and majors, the advising program was designed to improve students' college applications without altering their preferences.

The "business as usual" lighter-touch advising. I also developed a "business as usual" intervention, serving as a placebo test. I was interested in the following question: In the absence of this personalized advising, what for-profit consulting services would a student probably have access to? Between 2016 and 2017, I reviewed dozens of Chinese companies that were selling college application consulting services at a price range between 100 RMB and 500 RMB ($15-$80).[25] Excluding the obviously incorrect "application strategies," I kept a brief list of mostly harmless college application advice guidelines, which can be seen as a simplified, lighter-touch version of the college choice guidebook in 2016 (Ye, 2020). These tips were then provided to students in the "business as usual" group. A group of research assistants (not expert advisors) also answered generic questions. In many cases, students were directed to other online resources for further information. Research assistants did not answer any personalized questions about CEE score equating, short lists of colleges, or application planning.

[25]There were also more expensive personalized services charging thousands or tens of thousands RMB. But it is rare for students in the poorest regions to pay that much money.

3.3 A Student-Level Randomized Experiment

I conducted a large-scale randomized experiment in Ningxia to test the effectiveness of the machine learning-assisted personalized advising. The sample included the universe of Ningxia high school graduates who took the College Entrance Exam in 2017. I implemented student-level stratified randomization to increase statistical power. Using student information from the College Entrance Exam Registration Data, I generated randomization strata by school, track, gender, race, rural *hukou*, county of residence, and achievement. I classified high-achieving students using test score in the low-stakes graduation exam as the CEE score was not available at the time of randomization. Students who ranked above the 75th percentile in the high school graduation test - held in the fall semester of senior year - were classified as high-achieving students.

As shown in Figure 1, students were randomly assigned into one of the three groups: (1) 5,647 students were provided access to the machine learning-assisted personalized advising; (2) 5,370 students were provided access to the "business as usual" advising; and (3) the remaining 43,038 students served as the control group. Since targeting strategies may be more useful for high-achieving students (Tier 3 and Tier 4 colleges are not selective), I disproportionately (within randomization strata) assigned more high-achieving students to the machine learning advising group (45%; 28% in the "business as usual" group).

Implementation. When the CEE score report became available online and students could start to apply to college on June 21, the Ningxia Department of Education sent a text message to every student using the cellphone numbers from students' CEE registration records.[26] Immediately after that message, the Ningxia Department of Education sent another message to students in the two treatment groups, introducing the personalized advising opportunities provided by experts from Peking University and Ningxia University. Students were encouraged to contact the advisors using the online chat App. The text message was the same for the two intervention groups except for the contact information.[27] The research assistants verified each student's ID information and asked each

[26] All the students should register a cellphone number of their own or parents' cell phone. All the official notifications and information from the Department of Education are communicated using the registered cell phone number.

[27] For better coordination, we had five user account numbers, three for the machine learning group and two for the

student to complete a short online survey. Students were then directed to either a "machine learning" individual chat group or a "business as usual" individual chat group. Each chat group had three members: the treated student, one of the two expert advisors (an research assistant in the "business as usual" group), and an administrative assistant.

The machine learning-assisted personalized advising proceeded using three data analysis steps as described in Section 3.2. Students were provided with (1) automatic CEE score equating, (2) a short list of colleges, and (3) the predicted admissions probabilities of the colleges and majors they considered to apply to. The main task of the expert advisor was to help students check and fine-tune their candidate lists, and to provide more detailed guidance on specific questions. Most students kept in touch with inquiries and questions until they submitted their applications, and many of them informed us of their admissions results in July and August.

Interactions and conversations between the research assistant and the student in the "business as usual" advising were less frequent and much shorter than those in 2016 personalized advising program (Ye, 2020) and in the "machine learning" program. Students were provided with general guidelines and information about college applications. For example, students were provided explanations about the admissions mechanisms and they were suggested to apply to a mix of different types of colleges. However, the "business as usual" advising did not provide any detailed or personalized information about identifying and choosing specific reach, peer, and safety colleges nor how to place them in different orders. Students had to implement these strategies and make predictions on their own.

Summary statistics and validity. Table B.2 shows that student characteristics are well-balanced across groups given the student-level randomization. About 57% of students were from rural families; 32% were minorities; and more than 23% had repeated the 12th grade at least once. Compared with those who did not take the College Entrance Exam, this was a highly selected sample of "lucky" students who had overcome all the barriers from birth to grade 12.[28] Of this sample, 90% submitted college applications and 84% were admitted to college. Because the

"business as usual" group.

[28]Nationally, only about 40% of a birth cohort (18 million students) reach the stage of college application.

intervention randomization was independent of that for a parallel teacher intervention, there were 836 treated students in treated teachers' classes. Balance checks were still valid when excluding those overlapped students. Table B.3 confirms that student characteristics have statistically zero prediction power for the treatment status. All the joint F tests are statistically insignificant.

4 Results

4.1 Empirical Approach

To estimate the causal impact of the advising interventions - "machine learning" and "business as usual" - on college choice behaviors and admissions outcomes, I estimate an intent-to-treat effects (ITT) linear regression:

$$Y_i = \beta_0 + \beta_1 * T1(Machine\ learning)_i + \beta_2 * T2(Business\ as\ usual)_i + X_i * \gamma + \delta_s + \varepsilon_i \quad (2)$$

where Y_i is the outcome of interest for student i. $T1_i$ and $T2_i$ are indicator variables for student i receiving the text message invitations to the two advising groups (assignment to treatments), respectively. δ_s are strata fixed effects. All standard errors are clustered at the school level. I report joint test results for the two interventions, and test for the difference between the two: specifically, I test $H_0 : \beta_1 = \beta_2$.

X_i includes a student's CEE score to identify the "college choice" effect. In centralized systems, college admissions are jointly determined by a student's entrance exam score and her choice. Controlling for the entrance exam score in X_i, β_1 and β_2 estimates the treatment effects on a student's college access and match, through the impacts on her college choices. While treatment and control groups are well balanced in student demographics because of the stratified randomization (as shown in Table B.2 and Table B.3), X_i also controls for a vector demographics (gender, race, age, rural, track, repeater from the prior years) in the preferred specification to increase the statistical power and to reduce potential biasedness of effect size estimation. As expected, the results do not change if I exclude those student-level covariates.

Model (2) identifies the impacts of *being offered access to receive* personalized advising in college choice and application. I also estimate the treatment-on-the-treated effects (TOT) using a 2SLS regression, which measure the average effect of *receiving* the personalized advising on those who actually receive it. The first-stage regression examines the take-up of the two advising interventions:

$$\begin{aligned} Treated\ in\ T1_i =& \beta_0 + \beta_1 * T1_i + \beta_2 * T2_i + X_i * \gamma + \delta_s + \varepsilon_i \\ Treated\ in\ T2_i =& \beta_0 + \beta_1 * T1_i + \beta_2 * T2_i + X_i * \gamma + \delta_s + \varepsilon_i \end{aligned} \tag{3}$$

I then estimate TOT, the impacts of the exogenously-instrumented intervention participation ($\widehat{Treated\ in}\ T1_i$ and $\widehat{Treated\ in}\ T2$) on outcomes:

$$Y_i = \beta_0 + \beta_1 * \widehat{Treated\ in}\ T1_i + \beta_2 * \widehat{Treated\ in}\ T2_i + X_i * \gamma + \delta_s + \varepsilon_i \tag{4}$$

where the other specification issues are the same as in Model (2). ITT effects show the overall effects that we could expect if the program implementation is similar to what I did in 2017 when the take-up was low. TOT effects identify the potential intervention effects if we provide the machine learning assisted advising with stable program productivity to all, and all students take up their opportunities.[29]

I examine the primary college admissions outcomes as will be described in the following section, which include both extensive and intensive margins, because the interventions are primarily designed to improve college access and match. I also examine a list of exploratory measures of college choice behaviors. These measures are from the same domain and highly correlate with each other, so that the multiple hypothesis testing problem is minimal. Furthermore, I use the aggregated indices of each group of outcomes from a principal-component analysis. Additionally, I apply the method proposed by List, Shaikh and Xu (2016) to check the robustness of the results.

[29]This simple interpretation assumes that the average treatment effect is the identical to the average treatment effect on the treated, which is possible if students fully follow the machine learning based recommendations.

4.2 First stage results: Intervention Take-up

Table 2 reports the first stage regression results from Model (3), separately for the two advising interventions: machine learning assisted advising and "business as usual" advising. Column (1) shows that, on average, 3.6 percent of students (210 out of 5,647) who were provided the machine learning advising invitations eventually received the personalized assistance.[30] Column (5) shows that on average, 2.4 percent of students (134 out of 5,370) who were provided the "business as usual" advising invitations eventually received the low-touch personalized assistance. As expected, the results do not change when controlling for student-level covariates or the two random assignment indicators ($T1$ and $T2$) simultaneously. Columns (3) and (7) show that the parallel teacher incentives program did not impact the take-up rate. High-achieving students had slightly higher take-up (4.93% in $T1$ and 4.14% in $T2$; Chow test p-value = 0.209, which indicates no statistically significantly different take-up between the two groups). F statistics reject the null that the random assignment is a weak instrument for the actual take-up.

The take-up rate is somewhat surprisingly low, but higher than that in 2016 (1.5%). When designing the interventions, I planned for an expected take-up rate of 20% and prepared accordingly a team of expert advisers; thus I did not provide the conventional expert advising. One reason for this low take-up is the verification process (Alatas et al., 2016).[31] For the research purpose of identifying the treated students, student exam IDs and school IDs were used to verify and screen the targeted students. That is, I purposely denied most of the would-be always-takers from the control group. We also asked treated students to complete a 10-minute survey before the expert advisor provided advising services. As Hoxby and Turner (2013) suggest, students and parents may often be suspicious of this verification process, even though we only asked for their exam ID and school ID that could not be personally identified without the administrative data from the Department of

[30] We also provided advising to 4 "always taker" students (3 in $T1$) during the last few days.

[31] The other reason may be that, although nearly all of the students need assistance, the actual demand for the personalized advising may be still low. In 2018, I worked with collaborators to fully communicate with students about the benefits and the process of the advising using mails. Take-up increased to 10 percent. Hyman (2019) conducted a mail-based intervention that encouraged high-achieving high school seniors in Michigan to navigate a college information website. Its take-up is 9.8 percent.

Education. Over 1,800 users (students or parents) added the contact accounts as friends, accounting for 16 percent of the randomly assigned treated students (assuming few non-compliers). However, we finally provided college application advising to 347 high school graduates in Ningxia in the 2017 program. While this low take-up does not impact the internal validity of the estimates, it results in limited statistical power (e.g., significance tests among high achieving students) and an inability to infer the effect heterogeneities on a large scale.

4.3 Effects of Machine Learning Assisted Advising

Similar to the conventional personalized advising in the previous year, I find that the personalized advising program with the assistance of machine learning and related data analytics substantially improved college access - admission to a college, and match - the quality of an admitted college. Table 3 reports both the intent-to-treat effects (ITT) and the treatment-on-the-treated effects (TOT).

The first column shows that the machine learning assisted advising substantially increased students' college access. The TOT result suggests that, controlling for CEE score, demographics, and strata fixed effects, treated students had on average a 24.4 percentage points (pp) increase in their probability of college admissions, statistically significant at the 0.1 level. Note that the statistical power is limited by the low take-up. Accounting for the first-stage participation rate, the average effects (ITT) of offering the personalized advising program increased the college admission probability by 1 pp. This increase was from both increased application (column 2: TOT=12.8 pp, $p>0.1$) and improved college choice behavior (admission conditional on application). Columns (3) and (4) suggest that the increased admissions entirely shifted students from repeating another year - and retaking the CEE in 2018 - to on-time college enrollment in 2017. Students chose to retake the CEE after one year primarily because they were not satisfied with their CEE performance or admission offers. The personalized intervention helped students consider the best possible options conditional on their CEE scores and nudged students to enroll in college on time.

On the intensive margin, the personalized advising also substantively improved college match, which is measured by the quality of a student's admitted college. Column (5) of Table 3 examines the impacts on the college match index, which summarizes a family of five college quality measures using factor analysis. On average, treated students were admitted to colleges with 0.598 standard deviations higher quality as measured by the single index, statistically significant at the 0.05 level. The corresponding ITT effect of providing access to personalized advising is 0.022 s.d. ($p<0.05$). Given that college admissions are solely determined by a student's CEE score and her applications, the results suggest that in the counterfactual situation without receiving the machine learning assisted advising, a treated student had to increase her CEE score by 0.598 s.d. to be admitted to the same college. Comparing with many possible inputs in K-12 education, this paper presents a very effective, and relatively low cost, behaviorally-designed intervention to improve college access and match for low-income students.

The college match index consists of both contemporaneous college admissions scores (median, mean, minimum) and static (in the short term) national college ranking measures (standardized score and ranking percentile). The national college quality measures were constructed using college admissions data in all provinces from 1996-2017, as well as administrative data on institutional resources for every college in China. I use the national measures to minimize the potential bias of using admissions results from the same within-province cohort to denote college quality, for example, a college with few admission quotas occasionally admitting high-achieving students. The static measures also enable us to compare estimates across years.[32]

In column (6) of Table 3, I examine the impacts on the national college quality measure. Results are similar in that the personalized advising assisted the treated students to be admitted to colleges with a 0.77 s.d. higher quality in the national college ranking. Table B.4 presents very consistent results using the other four itemized college match outcomes. Column (7) excludes about 10,000 students who were not admitted to any college and presents the underestimated effects of "machine learning" (overestimates for "business as usual" effects). The point estimates remain

[32]The college peer median/mean CEE scores are very likely to be different for the same college in 2016 and 2017 depending on its applicant pools and admissions quota, as well as the CEE score distributions. All of the three statistics vary greatly across years. In this paper, the national college ranking data are the same for the same college in 2016 and 2017, providing a more stable measure of college quality.

large (TOT=0.262 s.d.), but are not precise enough to be statistically significant. Lastly, column (8) uses the dichotomous measure of undermatch using a cutoff score of 0.25 s.d., and shows that the machine learning advising program reduced undermatch by 28.8 percentage points. In magnitude, this equals the control group mean, or more than twice of the rural-urban gap.

4.4 How Did Machine Learning Work?

I designed the college application guide and advising to improve a student's college choice behaviors. Using the unique data on students' full college application lists, I test whether the improved college admissions outcomes stem from their application behavior changes. The construction of the (partial) strategy and preference measures is discussed in detail in Appendix Subsection C.1. I find compelling evidence that the machine learning advising nudged the treated students to use the correct data-based prediction strategies to apply for match colleges.

Table 4 reports both ITT and TOT effects on college choice behaviors. Panel A summarizes students' applications using the prediction information and shows that treated students in the machine learning-assisted advising group applied to colleges with higher expected admissions outcomes. Column (1) compares the *ex ante* quality of each student's applied colleges, measured by mean admissions scores in the previous years. Machine learning-assisted advising helped students apply to higher quality colleges. While column (2) suggests that applying to those higher quality colleges mechanically reduces average admissions probabilities, column (3) shows that there is still a large increase in the expected admissions outcome (product of college quality and admissions probability). The magnitude of the expected admissions outcome is close to the actual admissions outcome as reported in column (6) of Table 3. Using the *ex post* admissions scores in 2017 in column (4) yields similar results.

Panel B focuses on college choice strategies. Column (5) shows that, students who received the machine learning-assisted advising were 30.6 percentage points ($p<0.05$) more likely to apply to at least three colleges in the recommendation list, compared with that 30.5 percent of the control group students who applied to at least three colleges in the list. The list includes all the colleges that we ever recommended to any treated students with an estimated admission probability larger than 35%. This result clearly demonstrates that treated students followed our advising. In contrast, "business as usual" advising did not cause students to be more likely to apply to colleges in the list.

Column (6) uses the single principal-component factor index to summarize the effects on college choice behaviors. The machine learning advising largely and statistically significantly (at the 0.1 level) affected students' college choices. The next six columns present detailed results for each strategy and preference category (factor indices from a series of items). Results show that the effects of machine learning advising concentrated on improving students' targeting and general nudge strategies in their college choices. This finding is consistent with the descriptive results in Table 1 that these two groups of strategies are the most important factors driving college match, and is also consistent with the focus of the individualized advising intervention design. In Table B.5, I report the itemized results for the targeting and general nudge strategies. Results show that the improvement was not in a few occasional measures, but was universal across the domain of "good applications."[33] As a placebo test, the "business as usual" intervention shaped students' applications in the opposite direction: The treated students were less likely to use the optimal strategies based on data analytics.

Importantly, the personalized advising was designed only to improve students' strategic decision-making in college choice based on accurate predictions. It should not trade student preferences for the improvement in academic match.[34] Results show that neither the "machine learning" nor the "business as usual" interventions impacted student preferences, which confirms that the personalized advising using machine learning predictions worked through improving application strategies without impacting individual preferences for colleges and majors.

Why did "business as usual" not work? In stark contrast, the "business as usual" advising, as a placebo test that mimics a generic guide on the human expertise of college choice, had a

[33]For instance, students who received the machine learning assisted advising were much more likely to apply to academically matched colleges, and to list colleges in the correct descending order. They were also more likely to apply to a sufficient number of colleges and majors.

[34]Students had strongly motivated beliefs and preferences. A student survey in one Ningxia high school before the CEE in 2018 (N=1,190) shows that major preference, labor market prospects, college quality and cost are the main factors affecting students' initial college choices.

zero or even negative impact on college admissions results. For the primary outcomes of college access and match (columns 1 and 5), the joint equality tests reject the null hypothesis that the two individualized advising interventions had the same effects. This finding is consistent with the findings in Oreopoulos and Ford (2016), that decreased guidance in choosing eligible programs would limit the effectiveness of advising programs. One reason for the possible negative impact of "business as causal" intervention is that students may make their choices of repeating the 12th grade based on general advice such as "repeating increases college opportunities." In contrast, the machine learning-assisted advising nudged students to consider all the possible good college opportunities before deciding to repeat the 12th grade.[35] This argument is supported by the results in Table 3. The "machine learning" advising largely decreased repeating and accordingly increased on-time college enrollment. The "business as usual" had the opposite effect. Students who received the "business as usual" advising were more likely to repeat the 12th grade and less likely to enroll at a college in 2017. Excluding those who were not admitted to any college, column (7) shows that "business as usual" advising had a small and insignificant positive effect on college match. However, this estimate was biased upward.

Thinking fast? Slow! Most of the existing literature on college-going interventions emphasizes nudging students to take required actions on time to meet the application deadlines. I turn to the other aspect of time use in decision-making: haste makes waste. The seminal work by Kahneman (2011) describes two systems of human thinking - System 1 (thinking fast), and System 2 (thinking slow). System 1 forms automatic, first impressions of decision-making without deliberation, while System 2 involves problem-solving, analytical, and critical thinking. High school graduates may rely on their daily routine of System 1 thinking. The primary task of preparing for the college entrance exam, particularly in China, is to help students practice as much as possible and to train them for fast thinking during the test. Behavioral problems arise when students make

[35]It is arguable whether repeating is a good strategy. Goodman, Gurantz and Smith (2020) show that retaking the SAT improves admissions-relevant SAT scores. But Chinese students have to spend a whole year before retaking the CEE. In this paper, I define the not on-time enrollment as undermatch because too many students make repeating decisions without thoughtful considerations about college choices. This means that, presumably for some students it is optimal to retake the CEE, but the number of students who actually retake the CEE is much larger.

their college choices without thinking enough about their college options, resulting in not making a good college choice.

Using the sample of untreated students, Table B.6 reports estimates from four different strategies: (1) OLS without school fixed effects; (2) OLS with school fixed effects; (3) inverse-probability-weighted regression adjustment; and (4) IV using the random assignment to the two advising interventions as instrumental variables.[36] Results consistently show that spending more time making college choices is strongly positively correlated with better college access and match. The accordingly improved college application behaviors, especially in data-based targeting strategies, suggest that the "slow" students made a conscious effort in college choice.

The intervention design in this paper aimed to promote students' System 2 thinking and improve college match through effortful data-based predictions and targeting strategies. Figure A.4 shows the distribution of application time for students who were eligible to apply for selective colleges in 2017. Students in the machine learning-assisted advising group spent more time in their college choice decisions. In Table B.7, I formally test whether the personalized advising slowed down students. The first column shows the average ITT effects. Machine learning, on average, increased application time by 1 hour. The rescaled TOT effect is 27 hours ($p<0.05$). The effects concentrate on rural, male, non-minority, and low-achieving students. This finding is consistent with the results in Table B.8 that those students benefited from the machine learning assisted advising more than others. In contrast, I find that the "business as usual" intervention decreased students' decision time (ITT=-0.777; TOT=-32), which is consistent with the suggestive explanation that the brief guidelines in "business as usual" may have reduced the intrinsic motivation of students, and may have made them less likely to form thoughtful college choices.

[36]The IV estimates violate the exclusion restrictions because the interventions should have impacted other student behaviors in college choice as well as the time spent on navigating the online college applications. The estimates may overstate the treatment effects, which only serve as a comparison reference.

4.5 Heterogeneity Analysis: Who Benefits Most from the Personalized Advising?

I investigate the heterogeneous treatment effects in order to understand how the personalized advising would close the socioeconomic gaps in college access and match. The classical parametric approach is to add interactions between the treatment indicators and the conditioning variables of interest to the linear model (2). Table B.8 reports the results for $T1(Machine\ learning)$ while none of the estimates for $T2(Business\ as\ usual)$ are statistically and economically significant. The first row suggests that there were no heterogeneous take-up rates across subgroups except that high achieving students were more likely to participate in the advising program. Results show that rural, male, non-minority, and low-achieving students benefited more from the machine learning-assisted personalized advising. Those students were more likely to follow the machine learning recommendations or to use targeting strategies to apply to match colleges. Results using cluster-robust Causal Forests to uncover the heterogeneous advising treatment effects (Athey and Imbens, 2016; Wager and Athey, 2018; Athey et al., 2019) are qualitatively similar. Subsection C.4 provides further discussions.

4.6 Comparing the Effectiveness of Machine Learning and Human Instruction

I have shown robust evidence that machine learning-assisted advising substantially improves college access and match outcomes, through the combination of human expert instruction and machine learning predictions. A final analysis is to compare the effects of the "machine learning" approach in 2017 with the expert "human instruction" approach in 2016 in Ye (2020). Using the estimated effects of the expert human instruction in 2016 as a benchmark, I find that the machine learning approach had a similar impact on students' college access and match outcomes. For instance, the estimated effect of human instruction in 2016 on the college match outcome index is 0.210 s.d. for high achieving students. The machine learning assisted advising generates a 0.285 s.d. effect for high achieving students and a 0.598 s.d. for all students. Using national college ranking as a constant measure of college quality across years, both the machine learning approach and the expert human instruction approach largely shifted students to higher ranked colleges (by about 10 to 20 percentiles).

While the intervention effect was similar using either conventional expert advising or the machine learning assisted advising, machine learning greatly replaced human labors. There were six expert advisers who worked relentlessly to help 119 students in 2016 . But in 2017, only two expert advisors served 213 students. A simple calculation will show that, with the assistance of machine learning in simplifying the data analysis process, the efficiency and productivity of human instruction in the personalized advising program dramatically increased.[37] If we incorporate the administrative process into an online automatic system (e.g., artificial intelligence, Page and Gehlbach, 2017), the need for administrative assistants will be largely reduced (even to zero). To precisely quantify the increased productivity due to the introduction of machine learning remains as an open question for future research

5 Data vs. Algorithms: Understanding How Machine Learning Predictions Improve College Choice Decisions

This paper has so far presented compelling experimental evidence that machine learning predictions improve college access and match. Like many real-world applications, machine learning is only used as a "black box." This subsection aims to decode the black box and evaluate the relative importance of the key elements in machine learning predictions. A typical machine learning practice can be decomposed into four steps: (1) understanding domain knowledge, (2) preparing and pre-processing data, (3) training the learning models, and (4) applying results to guide decisions. During the college application season in August 2020, I conducted a survey experiment with 2,542 Chinese high school graduates to examine the impacts of differential access to various machine learning elements on college choice quality.

To measure college choice quality, I designed a college application competition similar to

[37] The advising productivity increased more than four times with the assistance of machine learning (= (213 students/2 counselors) - (119 students/6 counselors))/(119 students/6 counselors)). In some cases, we had a few more advisors to help with questions. There should still be a large improvement even when we very conservatively account for the inputs of these additional advisors.

the Kaggle data science challenges that asked participants to submit college applications for the same hypothetical applicant.[38] Compared with using a field experimental design, this approach had two advantages. First, it did not affect each participant's real life outcomes. For the concern of truth-telling, I used a set of award incentives to encourage participants to submit deliberate applications. Second, the competition only asked participants to consider college quality measured as the national ranking used in the previous section. This setup helped identify the impacts of college choice behaviors by ruling out the impacts of college and major preferences.

With the assistance of China Center for Education and Human Resources, I recruited online survey participants through various channels. The final sample included 2,542 high school graduates from all over the country (Figure A.5). However, due to the take-up differences, the sample was not nationally representative. For example, more than 48% of the survey participants had a parent with college education (vs. a 22% average in a national college student survey). Since disadvantaged students are more likely to benefit from machine learning-assisted advising, effects estimated in this survey experiment would serve as lower bound estimates. Following completion of consent form and a short background survey, participants read the introduction and rules for the "College Application Competition Against AI." The competition proceeded as follows.

1. *Information.* Participants were presented with a list of the name and national ranking of 195 colleges that had allocated their admissions quotas to non-STEM rank students in Chongqing in 2020 (the randomly selected context for the competition). They were also presented with an hypothetical applicant with a college entrance exam score of 566.

2. *Submission.* Participants submitted their college applications for the hypothetical applicant before the 2020 college admissions outcomes were released in mid-August. Each participant was required to apply to six colleges in a ranked order.

3. *"AI reference."* I pre-registered an application portfolio of six colleges that was generated by a machine learning optimal decision algorithm.

[38]The design idea of using hypothetical applicant was motivated by the Standardized Patient program in medical education that evaluates medical students in a simulated clinical environment.

4. *Simulating results.* After the 2020 college admissions outcomes in Chongqing were officially unveiled, based on the Deferred Acceptance mechanism, I used college admissions cutoffs to simulate the admissions results for the participants. Among the colleges that the hypothetical having admissions scores equal to or lower than the hypothetical applicant's score, each participant was assigned to be "admitted" to the college at the highest ranked order in her application portfolio.

5. *Determining winners.* The top three participants "admitted" to the highest ranked colleges among all participants were the winners. Participants who had ranked top 25% based on admissions results received cash rewards. Additionally, all the participants who had admissions results better than the "AI reference" would receive another special award.

In the "College Application Competition Against AI," I first tested the importance of the two main inputs to machine learning: human expertise (or domain knowledge) and data. The intuition was that neither students nor machine learning could make correct predictions without a good understanding of how the college admissions mechanism works or access to appropriately pre-processed data. I designed four treatment groups: (1) **Human expertise** that provided the "business as usual" guidance on college application strategies, (2) **Data** that provide admissions scores data of the 195 colleges in 2019, (3) **Human expertise + Data** , and (4) **Human expertise + Data + Classification** that provided additional information about the reach/match/safety types of the 195 colleges based on machine learning predictions. Figure A.6 displays the online survey experiment screen shots with specific information for each group.

I randomized each participant independently into one of the five groups (one control group and four treatment groups) with a probability of 20% (Figure 1). The automatic randomization algorithm of the survey website resulted in that covariates were well balanced across groups (Table B.9). The following linear regression was used to estimate the treatment effects (β_1 to β_4), controlling for the covariates X_i listed in Table B.9:

$$Y_i = \beta_0 + \beta_1 * T1_i + \beta_2 * T2_i + \beta_3 * T3_i + \beta_4 * T4_i + X_i * \gamma + \varepsilon_i \quad (5)$$

Table 5 reports the regression results. I constructed the outcome measures using the survey participants' submitted college applications and their simulated outcomes. Panel A focuses on college choice behaviors. On average, students in the control group applied to colleges with an average admission probability of 25.4% and treated students applied to colleges with higher admissions probabilities. Similar to the null effects of "business as usual" advising in the previous field experiment, access to human expertise ($T1$) did not statistically and substantially significantly improve college choices. Access to historical admissions data even without perfectly understanding the college admissions mechanisms ($T2$) largely helped students apply to colleges with higher admissions chances. Providing students with both the human expertise and data ($T3$) had the largest effects while providing additional classification information ($T4$) slightly made students more likely to apply to reach colleges. Columns 2-5 examined each survey participant's college application portfolio. On average, students applied to 4.4 "out of reach" colleges that had very low predicted admissions probability (based on the machine learning model I built) for the hypothetical applicant. Treated students reduced the number of "too high" colleges and increased the number of colleges that more closely matched the hypothetical applicant's college entrance exam score.

Panel B examines whether the improved college choice behaviors translated to increased college access and match. Following the previous field experiment, the primary college admissions outcomes included whether the hypothetical applicant was admitted to any college (column 6) and the quality indicated by admissions scores of the admitted college (column 9). Results are consistent with those in Panel A. Understanding the admissions rules and applications tips did not change admissions outcomes but access to data mattered. Access to both human expertise and data increased college admissions by more than 21 percentage points (vs. control mean of 45.1%) and admitted college quality by a 0.32 standard deviation (or 66 places in the national college ranking). Providing additional classification information did not statistically significantly change admissions outcomes. These large treatment effects speak to the machine learning-assisted advising effects I found in the previous field experiment.[39] If college admissions were not constrained to at most

[39]The control mean of college admissions is lower than it was expected to be (e.g., about 80% of students were admitted to college each year in Ningxia). This is because the low-stakes award competition motivated students to take

one offer per student, students in the control group would receive 1.4 offers while treated students (except in $T1$) would have about one more offer. As a result of these improvements, column 10 shows that treated students had much higher rankings in the AI competition. These results confirm that the advising effects reported in the 2017 RCT should not be driven by the low take-up rate.

In Figure 4, I further examine the heterogeneous treatment effects by each participant's predicted probabilities based on their background characteristics. The admissions probabilities were estimated from a random forest model built on student covariates in the survey and the control group sample (training data). Consistent with the findings in the previous experiment, participants who had lower predicted admissions chances (e.g., without college-educated parents, did not understand the Deferred Acceptance mechanism) would benefit more from the machine learning elements. One important finding is that, for those students, providing guidance about applications strategies also meaningfully improved their college access and match outcomes.

I also designed to test the effects of optimal decision algorithms by comparing the participants' "human performance" with the pre-registered "AI reference" application portfolio. The intuition was that machine learning might outperform human beings in the complex optimization process even under the same information set of domain knowledge and data. In the case of college choice, even if a student has perfect information to make accurate predictions of admissions chances for each single college, it becomes a combinatorial problem when considering to choose six out of 195 colleges and rank them in order. I built an algorithm to simplify the optimal college choice decisions by automatically searching for the best candidates based on pre-specified features (ranking, admissions probabilities) to maximize the expected outcome.

Results suggest that this algorithm had the best admission outcome. When considering the admitted college quality, the pre-registered "AI reference" was admitted to the college with the highest ranking among all the colleges that the hypothetical applicant was qualified for. Six of the 2,542 participants had the same outcome (1 in Control, 1 in $T1$, 2 in $T3$, 2 in $T4$). However, when considering multiple admissions offers, "AI reference" beat all the participants: "AI reference" had

more risks for"out of reach" colleges.

four admissions offers from the four highest ranked colleges among all the possible admissions. These results imply that, while machine learning algorithms may not outperform simpler models for simple prediction problems (Bird et al., 2020), they have the potentials to help students search for and make optimal decisions in the complex college choice decisions.

6 Conclusion

In this paper, I have asked how to effectively and efficiently improve college access and match at scale. Building on the new evidence that precise predictions of college admissions probabilities are the key to an optimal college choice decisions, I designed a personalized advising program to guide students to use data-based prediction strategies. Personalized advising substantially improved college admissions outcomes through affecting a student's college application behaviors without changing their preferences for colleges and majors. I proposed and tested a novel policy solution to scale up the labor-intensive personalized advising by increasing advising efficiency using big data-based machine learning algorithms.

Results show that machine learning-assisted advising achieved similar treatment effects to the conventional expert advising but largely reduced the time that an expert advisor used to serve a student. Disadvantaged students who need the support most were predicted to benefit more than their advantaged peers from the personalized advising. A supplemental survey experiment further confirmed that the combination of human expertise and machine learning methods in behaviorally motivated interventions has a high potential to improve personalized decisions and outcomes in education and aspects of life, particularly for those students who need the support most.

Data- and technology-based methods are increasingly useful for policy prediction problems. These new methods, implemented jointly with expertise and data, have wide application prospects in increasing intervention effectiveness in the pathways to and through college. Future work can joint this line of literature to test how to combine "human instruction" and "machine learning" to make the best possible decisions at scale (Castleman, Owen and Page, 2015; Page and Gehlbach, 2017; Miller and Skimmyhorn, 2018; Dynarski et al., 2021; Arteaga et al., 2021).

References

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

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

Acharya, Anal, and Devadatta Sinha. 2014. "Early prediction of students performance using machine learning techniques." *International Journal of Computer Applications*, 107(1).

Alatas, Vivi, Ririn Purnamasari, Matthew Wai-Poi, Abhijit Banerjee, Benjamin A Olken, and Rema Hanna. 2016. "Self-targeting: Evidence from a field experiment in Indonesia." *Journal of Political Economy*, 124(2): 371–427.

Arslan, Hayri Alper. 2018. "Preference estimation in centralized college admissions from reported lists." *Working paper*.

Arteaga, Felipe, Adam J Kapor, Christopher A Neilson, and Seth D Zimmerman. 2021. "Smart Matching Platforms and Heterogeneous Beliefs in Centralized School Choice." *NBER Working Paper No. 28946*.

Athey, Susan, and Guido Imbens. 2016. "Recursive partitioning for heterogeneous causal effects." *Proceedings of the National Academy of Sciences*, 113(27): 7353–7360.

Athey, Susan, Julie Tibshirani, Stefan Wager, et al. 2019. "Generalized random forests." *The Annals of Statistics*, 47(2): 1148–1178.

Aulck, Lovenoor, Nishant Velagapudi, Joshua Blumenstock, and Jevin West. 2016. "Predicting student dropout in higher education." *arXiv preprint arXiv:1606.06364*.

Avery, Christopher, Benjamin L Castleman, Michael Hurwitz, Bridget T Long, and Lindsay C Page. 2020. "Digital messaging to improve college enrollment and success." *NBER*

Working Paper No. 27897.

Balinski, Michel, and Tayfun Sönmez. 1999. "A tale of two mechanisms: student placement." *Journal of Economic theory*, 84(1): 73–94.

Bergman, Peter, Jeffrey T Denning, and Dayanand Manoli. 2019. "Is information enough? The effect of information about education tax benefits on student outcomes." *Journal of Policy Analysis and Management*, 38(3): 706–731.

Bettinger, Eric P, and Brent J Evans. 2019. "College guidance for all: A randomized experiment in pre-college advising." *Journal of Policy Analysis and Management*, 38(3): 579–599.

Bettinger, Eric P, Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu. 2012. "The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment." *The Quarterly Journal of Economics*, 127(3): 1205–1242.

Bird, Kelli A, Benjamin L Castleman, Jeffrey T Denning, Joshua Goodman, Cait Lamberton, and Kelly Ochs Rosinger. 2021. "Nudging at scale: Experimental evidence from FAFSA completion campaigns." *Journal of Economic Behavior & Organization*, 183: 105–128.

Bird, Kelli A, Benjamin L Castleman, Zachary Mabel, and Yifeng Song. 2020. "Worth its weight in gold? A systematic comparison of predictive analytics methods in higher education." *Working Paper*.

Black, Sandra E, Kalena E Cortes, and Jane Arnold Lincove. 2020. "Apply yourself: Racial and ethnic differences in college application." *Education Finance and Policy*, 15(2): 209–240.

Bowen, William G, Matthew M Chingos, and Michael S McPherson. 2009. *Crossing the finish line: Completing college at America's public universities.* Princeton University Press.

Bulman, George. 2015. "The effect of access to college assessments on enrollment and attainment." *American Economic Journal: Applied Economics*, 7(4): 1–36.

Burch, Patricia, Annalee Good, and Carolyn Heinrich. 2016. "Improving access to, quality, and the effectiveness of digital tutoring in K–12 education." *Educational Evaluation and Policy Analysis*, 38(1): 65–87.

Burkhardt, JC, SL DesJardins, CA Teener, SE Gay, and SA Santen. 2018. "Predicting Medical School Enrollment Behavior: Comparing an Enrollment Management Model to Expert Human Judgment." *Academic Medicine: Journal of the Association of American Medical Colleges*, 93(11S): S68–S73.

Campbell, Stuart, Lindsey Macmillan, Richard Murphy, and Gillian Wyness. 2021. "Matching in the Dark? Inequalities in student to degree match." National Bureau of Economic Research.

Carrell, Scott, and Bruce Sacerdote. 2017. "Why do college-going interventions work?" *American Economic Journal: Applied Economics*, 9(3): 124–51.

Carruthers, Celeste K, and William F Fox. 2016. "Aid for all: College coaching, financial aid, and post-secondary persistence in Tennessee." *Economics of Education review*, 51: 97–112.

Castleman, Benjamin. 2017. "Behavioral insights for federal higher education policy." *Urban Institute*.

Castleman, Benjamin, and Joshua Goodman. 2018. "Intensive college counseling and the enrollment and persistence of low-income students." *Education Finance and Policy*, 13(1): 19–41.

Castleman, Benjamin L, Laura Owen, and Lindsay C Page. 2015. "Stay late or start early? Experimental evidence on the benefits of college matriculation support from high schools versus colleges." *Economics of Education Review*, 47: 168–179.

Castleman, Benjamin L, Saul Schwartz, and Sandy Baum. 2015. *Decision making for student success: Behavioral insights to improve college access and persistence.* Routledge.

Chalfin, Aaron, Oren Danieli, Andrew Hillis, Zubin Jelveh, Michael Luca, Jens Ludwig, and Sendhil Mullainathan. 2016. "Productivity and selection of human capital with machine learning." *American Economic Review*, 106(5): 124–27.

Chen, Tianqi, and Carlos Guestrin. 2016. "Xgboost: A scalable tree boosting system." 785–794.

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

Damgaard, Mette Trier, and Helena Skyt Nielsen. 2018. "Nudging in education." *Economics of Education Review*, 64: 313–342.

Daniel, Ben. 2015. "Big Data and analytics in higher education: Opportunities and challenges." *British Journal of Educational Technology*, 46(5): 904–920.

Davis, Jonathan, Jonathan Guryan, Kelly Hallberg, and Jens Ludwig. 2017. "The economics of scale-up." *NBER Working Paper No. W23925*.

Dillon, Eleanor Wiske, and Jeffrey Andrew Smith. 2017. "The determinants of mismatch between students and colleges." *Journal of Labor Economics*, 35(1): 45–66.

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

Dinkelman, Taryn, and Claudia Martínez A. 2014. "Investing in schooling in Chile: The role of information about financial aid for higher education." *Review of Economics and Statistics*, 96(2): 244–257.

Dynarski, Susan, and Judith Scott-Clayton. 2013. "Financial aid policy: Lessons from research." *The Future of Children*, 23(1): 67–91.

Dynarski, Susan, CJ Libassi, Katherine Michelmore, and Stephanie Owen. 2021. "Closing the gap: The effect of reducing complexity and uncertainty in college pricing on the choices of low-income students." *American Economic Review*.

Evans, Brent J, and Angela Boatman. 2019. "Understanding how Information Affects Loan Aversion: A Randomized Control Trial of Providing Federal Loan Information to High School Seniors." *The Journal of Higher Education*, 1–33.

Farrell, Max H, Tengyuan Liang, and Sanjog Misra. 2018. "Deep neural networks for estimation and inference: Application to causal effects and other semiparametric estimands." *arXiv preprint arXiv:1809.09953*.

Fesler, Lily. 2020. "Opening the Black Box of College Counseling." *CEPA Working Paper No. 20-03*.

French, Robert, and Philip Oreopoulos. 2017. "Behavioral barriers transitioning to college." *Labour Economics*, 47: 48–63.

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

González, Julie M Byers, and Stephen L DesJardins. 2002. "Artificial neural networks: A new approach to predicting application behavior." *Research in Higher Education*, 43(2): 235–258.

Goodman, Joshua, and Melanie Rucinski. 2018. "Increasing diversity in Boston's Exam Schools." *Rappaport Institute for Greater Boston Policy Brief*.

Goodman, Joshua, Oded Gurantz, and Jonathan Smith. 2020. "Take two! SAT retaking and college enrollment gaps." *American Economic Journal: Economic Policy*, 12(2): 115–58.

Goodman, Sarena. 2016. "Learning from the test: Raising selective college enrollment by providing information." *Review of Economics and Statistics*, 98(4): 671–684.

Gurantz, Oded, Jessica Howell, Michael Hurwitz, Cassandra Larson, Matea Pender, and Brooke White. 2020*a*. "A national-level informational experiment to promote enrollment in selective colleges." *Journal of Policy Analysis and Management*.

Gurantz, Oded, Matea Pender, Zachary Mabel, Cassandra Larson, and Eric Bettinger. 2020*b*. "Virtual advising for high-achieving high school students." *Economics of Education Review*, 75: 101974.

Hafalir, Isa E, Rustamdjan Hakimov, Dorothea Kübler, and Morimitsu Kurino. 2018. "College admissions with entrance exams: Centralized versus decentralized." *Journal of Economic Theory*, 176: 886–934.

Hastings, Justine, Christopher A Neilson, and Seth D Zimmerman. 2018. "The effects of earnings disclosure on college enrollment decisions." *NBER Working Paper No. 21300.*

Heller, Sara B, Anuj K Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, and Harold A Pollack. 2017. "Thinking, fast and slow? Some field experiments to reduce crime and dropout in Chicago." *The Quarterly Journal of Economics*, 132(1): 1–54.

Herzog, Serge. 2006. "Estimating student retention and degree-completion time: Decision trees and neural networks vis-à-vis regression." *New directions for institutional research*, 2006(131): 17–33.

Hillman, Nicholas W. 2016. "Geography of college opportunity: The case of education deserts." *American Educational Research Journal*, 53(4): 987–1021.

Howell, Jessica S, and Matea Pender. 2016. "The costs and benefits of enrolling in an academically matched college." *Economics of Education Review*, 51: 152–168.

Hoxby, Caroline, and Christopher Avery. 2013. "The missing "one-offs": The hidden supply of high-achieving, low-income students." *Brookings Papers on Economic Activity*, 2013(1): 1–65.

Hoxby, Caroline M. 2000. "The effects of geographic integration and increasing competition in the market for college education." *NBER Working Paper 6323*.

Hoxby, Caroline M, and Sarah Turner. 2013. "Expanding college opportunities for high-achieving, low-Income students." *SIEPR Discussion Paper No. 12-014*.

Hurwitz, Michael, Jonathan Smith, Sunny Niu, and Jessica Howell. 2015. "The Maine question: How is 4-year college enrollment affected by mandatory college entrance exams?" *Educational Evaluation and Policy Analysis*, 37(1): 138–159.

Hurwitz, Michael, Preeya P Mbekeani, Margaret M Nipson, and Lindsay C Page. 2017. "Surprising ripple effects: How changing the SAT score-sending policy for low-income students impacts college access and success." *Educational Evaluation and Policy Analysis*, 39(1): 77–103.

Hyman, Joshua. 2017. "ACT for all: The effect of mandatory college entrance exams on postsecondary attainment and choice." *Education Finance and Policy*, 12(3): 281–311.

Hyman, Joshua. 2019. "Can Light-Touch College-Going Interventions Make a Difference? Evidence from a Statewide Experiment in Michigan." *Journal of Policy Analysis and Management*.

Imai, Kosuke, Marc Ratkovic, et al. 2013. "Estimating treatment effect heterogeneity in randomized program evaluation." *The Annals of Applied Statistics*, 7(1): 443–470.

Jacob, Brian, Brian McCall, and Kevin Stange. 2018. "College as Country Club: Do Colleges Cater to Students' Preferences for Consumption?" *Journal of Labor Economics*, 36(2): 309–348.

J-PAL. 2018. "Simplifying barriers along the bridge to college." *J-PAL Policy Bulletin*.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow.* London: Macmillan.

Kamada, Yuichiro, and Fuhito Kojima. 2015. "Efficient matching under distributional constraints: Theory and applications." *American Economic Review*, 105(1): 67–99.

Kane, Thomas J. 1999. *The price of admission: Rethinking how Americans pay for college.* Brookings Institution Press.

Kapor, Adam, Christopher A Neilson, and Seth D Zimmerman. 2020. "Heterogeneous beliefs and school choice mechanisms." *American Economic Review*, 110(5): 1274–1315.

Klasik, Daniel. 2013. "The ACT of enrollment: The college enrollment effects of state-required college entrance exam testing." *Educational researcher*, 42(3): 151–160.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2017. "Human decisions and machine predictions." *The Quarterly Journal of Economics*, 133(1): 237–293.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer. 2015. "Prediction policy problems." *American Economic Review*, 105(5): 491–95.

Lavecchia, Adam M, Heidi Liu, and Philip Oreopoulos. 2016. "Behavioral economics of education: Progress and possibilities." In *Handbook of the Economics of Education*. Vol. 5, 1–74. Elsevier.

Lechner, Michael, and Jeffrey Smith. 2007. "What is the value added by caseworkers?" *Labour economics*, 14(2): 135–151.

List, John A, Azeem M Shaikh, and Yang Xu. 2016. "Multiple hypothesis testing in experimental economics." *NBER Working Paper No. 21875*.

Long, Bridget Terry. 2004. "How have college decisions changed over time? An application of the conditional logistic choice model." *Journal of Econometrics*, 121(1): 271–296.

Loyalka, Prashant, Binzhen Wu, and Xiaoyang Ye. 2017. "Out of the gauntlet and into the fire: How college choices in centralized admissions systems disadvantage disadvantaged students." *Working paper*.

Machado, Cecilia, and Christiane Szerman. 2017. "Centralized admissions and the student-college match." *Working paper*.

Manski, Charles F, and David A Wise. 1983. *College choice in America.* Harvard University Press.

McKenzie, David, and Dario Sansone. 2017. "Man vs. machine in predicting successful entrepreneurs: Evidence from a business plan competition in Nigeria." *World Bank Policy Research Working Paper 8271*.

Miller, Brian J, and William L Skimmyhorn. 2018. "I want you! Expanding college access through targeted recruiting efforts." *Education Finance and Policy*, 13(3): 395–418.

Miller, Lois. 2017. "Reaching further: The role of distance in college undermatching." *https://scholarship.depauw.edu/studentresearch/72*.

Mulhern, Christine. 2021. "Changing college choices with personalized admissions information at scale: Evidence on Naviance." *Journal of Labor Economics*, 39(1): 219–262.

Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine learning: An applied econometric approach." *Journal of Economic Perspectives*, 31(2): 87–106.

Muralidharan, Karthik, Abhijeet Singh, and Alejandro J Ganimian. 2018. "Disrupting education? Experimental evidence on technology-aided instruction in India." *American Economic Review*, Forthcoming.

Muralidharan, Karthik, and Paul Niehaus. 2017. "Experimentation at scale." *Journal of Economic Perspectives*, 31(4): 103–24.

Neilson, Christopher. 2019. "The rise of centralized choice and assignment mechanisms in education markets around the world." *Working paper*.

Oreopoulos, Philip, and Reuben Ford. 2016. "Keeping college options open: A field experiment to help all high school seniors through the college application process." *NBER Working Paper 22320*.

Oreopoulos, Philip, and Uros Petronijevic. 2018. "Student coaching: How far can technology go?" *Journal of Human Resources*, 53(2): 299–329.

Oreopoulos, Philip, Robert S Brown, and Adam M Lavecchia. 2017. "Pathways to education: An integrated approach to helping at-risk high school students." *Journal of Political Economy*, 125(4): 947–984.

Ovink, Sarah, Demetra Kalogrides, Megan Nanney, and Patrick Delaney. 2018. "College match and undermatch: Assessing student preferences, college proximity, and inequality in post-college outcomes." *Research in Higher Education*, 59(5): 553–590.

Page, Lindsay C, and Hunter Gehlbach. 2017. "How an artificially intelligent virtual assistant helps students navigate the road to college." *AERA Open*, 3(4): 2332858417749220.

Page, Lindsay C, and Judith Scott-Clayton. 2016. "Improving college access in the United States: Barriers and policy responses." *Economics of Education Review*, 51: 4–22.

Page, Lindsay C, Stacy S Kehoe, Benjamin L Castleman, and Gumilang A Sahadewo. 2017. "More than dollars for scholars: The impact of the Dell Scholars Program on college access, persistence and degree attainment." *Journal of Human Resources*, 0516–7935r1.

Pallais, Amanda. 2015. "Small differences that matter: Mistakes in applying to college." *Journal of Labor Economics*, 33(2): 493–520.

Pathak, Parag A, and Tayfun Sönmez. 2013. "School admissions reform in Chicago and England: Comparing mechanisms by their vulnerability to manipulation." *American Economic Review*, 103(1): 80–106.

Perna, Laura W. 2006. "Studying college access and choice: A proposed conceptual model." In *Higher Education: Handbook of Theory and Research*. 99–157. Springer.

Peter, Frauke, C Katharina Spiess, and Vaishali Zambre. 2018. "Informing Students about College: An Efficient Way to Decrease the Socio-Economic Gap in Enrollment: Evidence from a Randomized Field Experiment." *DIW Berlin Discussion Paper No. 1770.*

Phillips, Meredith, and Sarah J Reber. 2019. "Does vrtual advising increase college enrollment? Evidence from a random assignment college access field experiment." National Bureau of Economic Research.

Powers, Scott, Junyang Qian, Kenneth Jung, Alejandro Schuler, Nigam H Shah, Trevor Hastie, and Robert Tibshirani. 2018. "Some methods for heterogeneous treatment effect estimation in high dimensions." *Statistics in Medicine*, 37(11): 1767–1787.

Schmidt, Rebecca A, and CJ Park. 2021. "Nonacademic interventions for postsecondary enrollment and success with rural and high-poverty populations: A systematic evidence review." *Arlington, VA: SRI International, Regional Educational Laboratory Appalachia.*

Smith, Jonathan, Matea Pender, and Jessica Howell. 2013. "The full extent of student-college academic undermatch." *Economics of Education Review*, 32: 247–261.

Taylor, Eric. 2018. "New technology and teacher productivity." *Working paper*.

Thaler, Richard H, and Cass R Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness.* Yale University Press.

Wager, Stefan, and Susan Athey. 2018. "Estimation and inference of heterogeneous treatment effects using random forests." *Journal of the American Statistical Association*, 113(523): 1228–1242.

White House. 2014. "Increasing college opportunity for low-income students: Promising models and a call to action." *Report of the Executive Office of the President*.

Ye, Xiaoyang. 2020. "Improving college choice for the poorest students: Results from randomized experiments in centralized admissions." *Working Paper*.

Figures

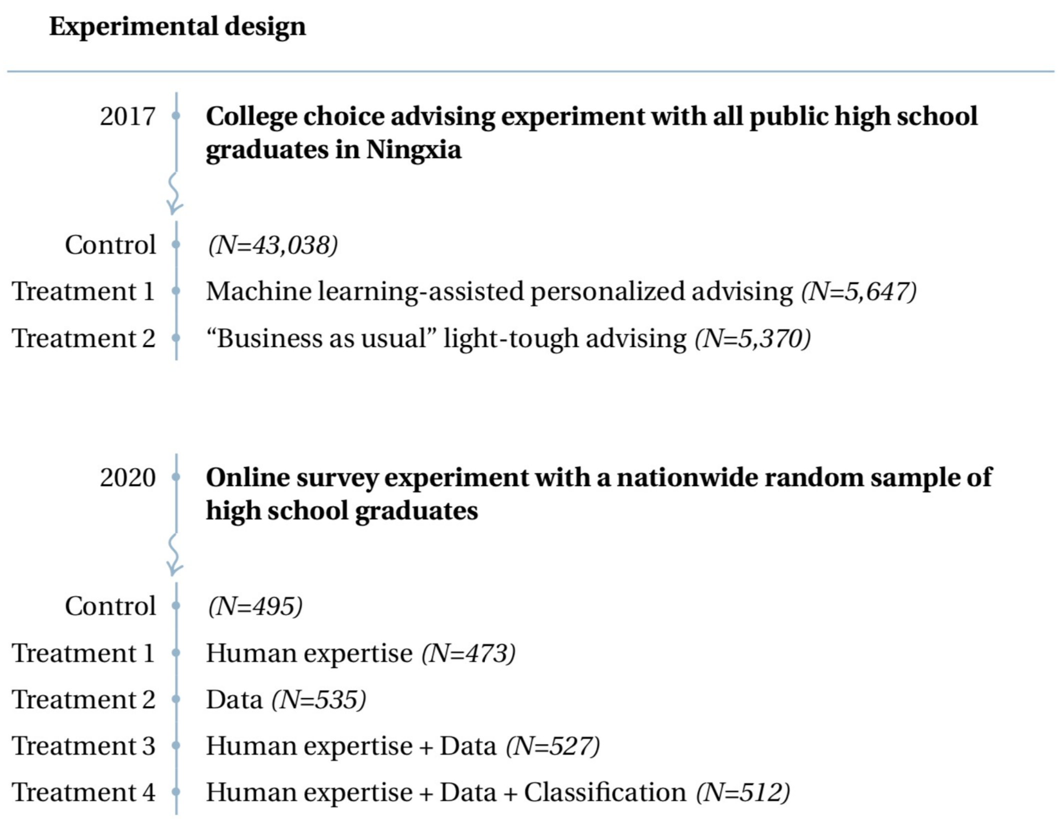
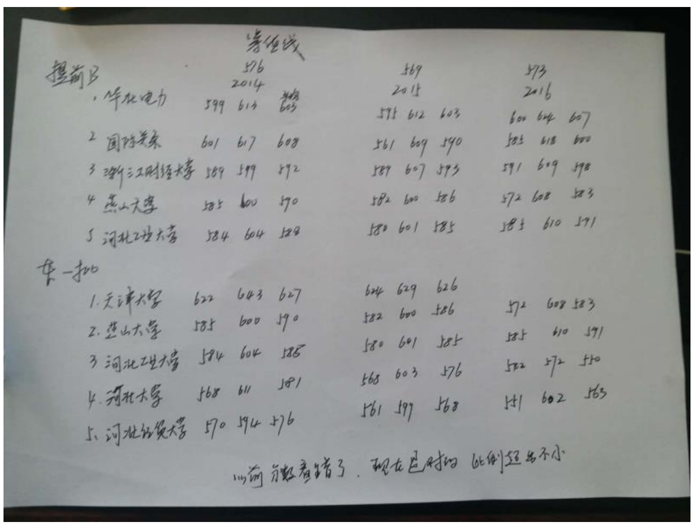
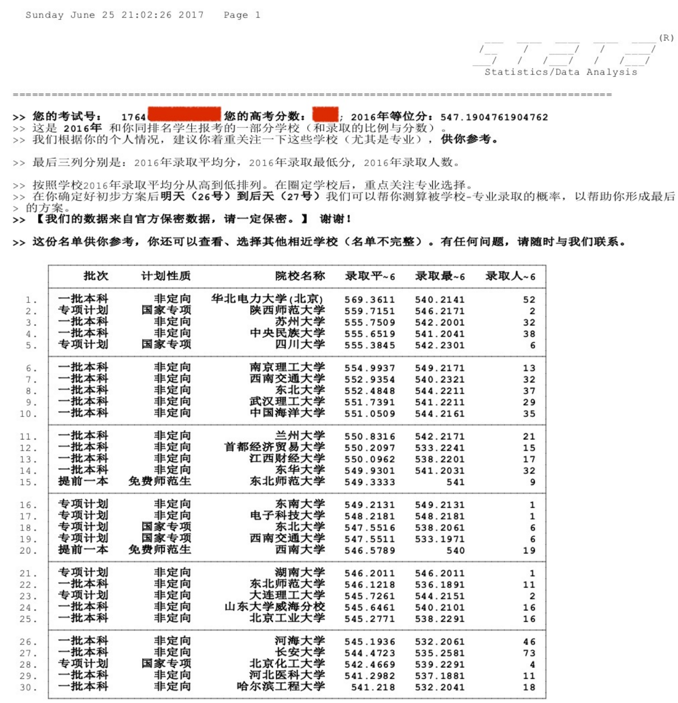


Figure 1: Experimental design

Notes: This figure shows the experimental design of the *Bright Future of China Project* in Ningxia in 2017 and the supplemental survey experiment in 2020. In the 2017 field experiment, randomization strata of student interventions were by school, track, gender, race, rural hukou, county of residence, and achievement (classifying high-achieving students using low-stakes graduation test scores). In the 2020 experiment, randomization was independently at the student level.



(a) Manual score equating



Sunday June 25 21:02:26 2017   Page 1

Statistics/Data Analysis

>> 您的考试号： 1764[redacted]您的高考分数：[redacted]; 2016年等位分：547.1904761904762
>> 这是 2016年 和你同排名学生报考的一部分学校（和录取的比例与分数）。
>> 我们根据你的个人情况，建议你着重关注一下这些学校（尤其是专业），供你参考。

>> 最后三列分别是：2016年录取平均分，2016年录取最低分，2016年录取人数。

>> 按照学校2016年录取平均分从高到低排列。在圈定学校后，重点关注专业选择。
>> 在你确定好初步方案后明天（26号）到后天（27号）我们可以帮你测算被学校-专业录取的概率，以帮助你形成最后
> 的方案。
>> 【我们的数据来自官方保密数据，请一定保密。】 谢谢！

>> 这份名单供你参考，你还可以查看、选择其他相近学校（名单不完整）。有任何问题，请随时与我们联系。

	批次	计划性质	院校名称	录取平~6	录取最~6	录取人~6
1.	一批本科	非定向	华北电力大学(北京)	569.3611	540.2141	52
2.	专项计划	国家专项	陕西师范大学	559.7151	546.2171	2
3.	一批本科	非定向	苏州大学	555.7509	542.2001	32
4.	一批本科	非定向	中央民族大学	555.6519	541.2041	38
5.	专项计划	国家专项	四川大学	555.3845	542.2301	6
6.	一批本科	非定向	南京理工大学	554.9937	549.2171	13
7.	一批本科	非定向	西南交通大学	552.9354	540.2321	32
8.	一批本科	非定向	东北大学	552.4848	544.2211	37
9.	一批本科	非定向	武汉理工大学	551.7391	541.2211	29
10.	一批本科	非定向	中国海洋大学	551.0509	544.2161	35
11.	一批本科	非定向	兰州大学	550.8316	542.2171	21
12.	一批本科	非定向	首都经济贸易大学	550.2097	533.2241	15
13.	一批本科	非定向	江西财经大学	550.0962	538.2201	17
14.	一批本科	非定向	东华大学	549.9301	541.2031	32
15.	提前一本	免费师范生	东北师范大学	549.3333	541	9
16.	专项计划	非定向	东南大学	549.2131	549.2131	1
17.	专项计划	非定向	电子科技大学	548.2181	548.2181	1
18.	专项计划	国家专项	东北大学	547.5516	538.2061	6
19.	专项计划	国家专项	西南交通大学	547.5511	533.1971	6
20.	提前一本	免费师范生	西南大学	546.5789	540	19
21.	专项计划	非定向	湖南大学	546.2011	546.2011	1
22.	一批本科	非定向	东北师范大学	546.1218	536.1891	11
23.	专项计划	非定向	大连理工大学	545.7261	544.2151	2
24.	一批本科	非定向	山东大学威海分校	545.6461	540.2101	16
25.	一批本科	非定向	北京工业大学	545.2771	538.2291	16
26.	一批本科	非定向	河海大学	545.1936	532.2061	46
27.	一批本科	非定向	长安大学	544.4723	535.2581	73
28.	专项计划	国家专项	北京化工大学	542.4669	539.2291	4
29.	一批本科	非定向	河北医科大学	541.2982	537.1881	11
30.	一批本科	非定向	哈尔滨工程大学	541.218	532.2041	18

(b) College short list in 2017

Figure 2: How does *Stata* reduce human labor?

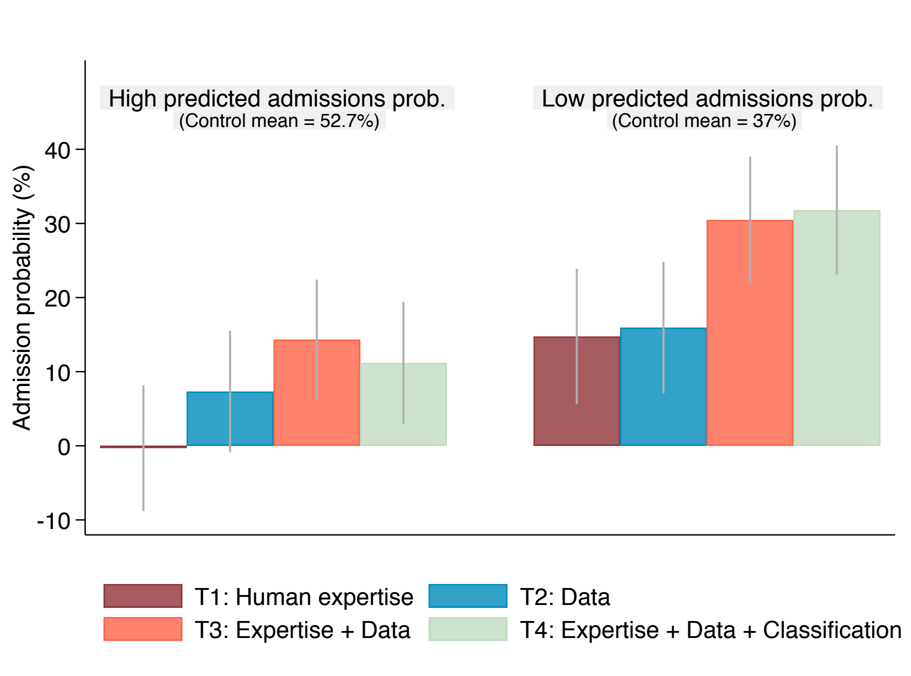
Notes: **Panel A** shows a score-equating table that one student completed (this is from a student in another province in 2017, which is similar to the individualized advising process in Ningxia in 2016). For a short list of two tiers (Tier 1 - Early admissions, Tier 1) and five colleges in each tier, she collected the admissions scores (maximum, mean, minimum) of each college for the past three years. On the top of the table, she listed her equated CEE scores in these three years. On the bottom, she noted that the previous table she returned to me had a mistake in the equated scores (then the comparisons were wrong). This table may have taken an hour or so (much longer if including the search time for the short listed colleges). It took much longer in the initial round of narrowing down the college options to a short list.
Panel B shows an automatic output of score-equating and college short list using *Stata*. It took several seconds after we typed in a student's ID. The *Stata* shortlist provided additional information like tier, special program, and admission quota. The length of the list was flexible upon a student's request.

(a) A low predicted probablity

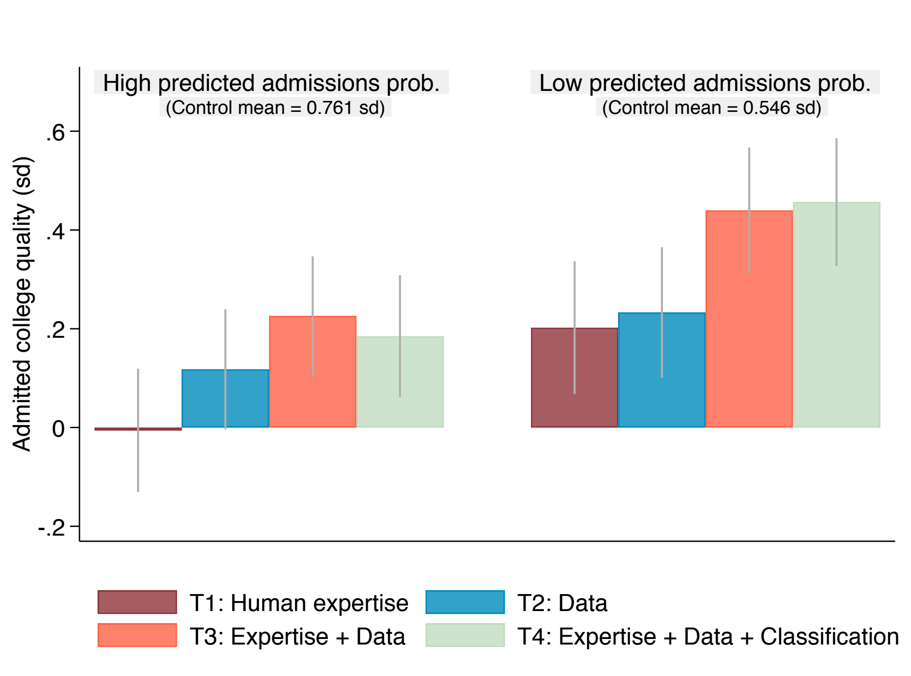
(b) A high predicted probability

Figure 3: How does machine learning work in 2017?

Notes: This figure shows the machine learning interface (designed by **Keqiang Li & Tzuyi Yu** at the University of Michigan) that we used in the 2017 fieldwork. We predicted the admissions probability for each college-major-rank list for each student. The right column shows the relevant information (match tier, college-level average admissions score, major-level average admissions score/quota in the prior year), most importantly, the predicted admission probability. We used different colors (red, blue, green) to indicate reach, peer and safety types. This information was used to assist the personalized advising. Advisers had access to the predicted probabilities of all the short-listed colleges and majors for each student. We shared the output pictures with students. In future work, this interface could be potentially hosted in a website for scale-up applications.



(a) Outcome = Admitted to any college



(b) Outcome = Admitted college quality

Figure 4: Heterogeneous effects of machine learning elements by predicted admissions probabilities

Notes: This figure plots the heterogeneous effects of machine learning elements by predicted admissions probabilities in the 2020 survey experiment. The admissions probabilities are estimated from a random forest model built on student covariates in the survey and the control group sample (training data). The vertical lines are 95% confidence intervals.

Tables

Table 1: College choices and the poverty gap in college match

	Mean		Outcome: Index of college match							
	Rural	Urban	Without school FE				With school FE			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Rural-urban gap (β_1)			-0.107***	-0.064***	-0.049***	-0.040***	-0.063***	-0.035***	-0.025***	-0.023***
			(0.011)	(0.009)	(0.008)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
(Strategies)										
Targeting	-0.159	0.155		0.204***	0.180***	0.207***		0.209***	0.182***	0.210***
				(0.008)	(0.009)	(0.010)		(0.008)	(0.009)	(0.010)
General nudge	-0.174	0.178			0.075***	0.070***			0.082***	0.073***
					(0.004)	(0.004)			(0.005)	(0.004)
(Preferences)										
Special programs	-0.011	-0.055				0.030***				0.029***
						(0.004)				(0.004)
Tuition & quota	0.048	-0.008				-0.064***				-0.064***
						(0.008)				(0.008)
Location	-0.300	0.287				0.010				0.020**
						(0.007)				(0.008)
Major	-0.050	0.072				-0.016***				-0.014***
						(0.003)				(0.003)
Observations			35,332	35,332	35,332	35,332	35,332	35,332	35,332	35,332
R-squared			0.713	0.747	0.751	0.756	0.719	0.751	0.754	0.758

Notes: This table reports the OLS regression (Model (1)) results for the partial correlations between college application behaviors and the college match index (standardized), using data from those who submitted college applications in the control group in 2017. Application behaviors are constructed using the full applications data, as described in Appendix Subsection C.1. Columns (1) and (2) present the mean values of each college behavior index for rural and urban students. Column (3) shows the rural-urban gap in college admissions using the full untreated sample (same as in column (3) of Panel C in Table B.1). The next three columns add the strategy and preference measures (principal component factor indices) stepwise. Columns (7)-(10) control for high school fixed effects. All regressions include a student's CEE score and other demographic covariates. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2: First stage: Take-up of individualized advising programs

	Treated in T1				Treated in T2			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
T1 (machine learning)	0.036***	0.036***	0.036***	0.036***				-0.001***
	(0.003)	(0.003)	(0.003)	(0.003)				(0.000)
T2 (business as usual)				-0.000**	0.024***	0.024***	0.024***	0.024***
				(0.000)	(0.002)	(0.002)	(0.002)	(0.002)
T1*Teacher incentives			0.001*					
			(0.001)					
T2*Teacher incentives							-0.001	
							(0.001)	
Teacher incentives			0.003				0.001	
			(0.009)				(0.008)	
Rural		0.000	0.000	0.000		-0.000	-0.000	-0.001
		(0.002)	(0.002)	(0.002)		(0.001)	(0.001)	(0.001)
Female		-0.000	-0.000	-0.000		-0.001	-0.001	-0.001*
		(0.001)	(0.001)	(0.001)		(0.001)	(0.001)	(0.001)
Minority		-0.000	-0.000	0.000		-0.001	-0.001	-0.001
		(0.003)	(0.003)	(0.003)		(0.003)	(0.003)	(0.001)
Age		0.002*	0.002*	0.002*		-0.002	-0.002	-0.001
		(0.001)	(0.001)	(0.001)		(0.001)	(0.001)	(0.001)
STEM		-0.000	-0.000	-0.000		-0.002	-0.002	-0.002*
		(0.001)	(0.001)	(0.001)		(0.001)	(0.001)	(0.001)
Repeater		-0.002**	-0.002**	-0.001**		0.000	0.000	0.001
		(0.001)	(0.001)	(0.001)		(0.001)	(0.001)	(0.001)
CEE score		0.002***	0.002***	0.002***		0.001***	0.001***	0.001***
		(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
F stat (excluded instruments)	141.4	142.0	75.8	72.8	134.5	134.2	68.3	68.2
Sanderson-Windmeijer F stat				190.8				154.4
Observations	48,685	48,685	48,685	54,055	48,408	48,408	48,408	54,055

Notes: This table reports the OLS regression (Model (3)) results of the take-up of the individualized advising interventions in 2017. Over 1,800 users (students or parents; about 16% of the treatment group size) added us as friends in the online message App (WeChat), but many of them refused to provide their exam ID and school ID for verification. Students in both groups received the same text message (the only exception is the contact information). Take-up rates among high-achieving students were 4.8% and 4.1% for the two treatment groups. All regressions control for strata fixed effects. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 3: ITT and TOT effects on college access and match of the individualized advising programs

	Admission (=1)		Application (=1)		Enrollment in 2017 (=1)		Repeating in 2018 (=1)	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1)		(2)		(3)		(4)	
Control mean	0.837		0.900		0.765		0.209	
Control sd	[0.369]		[0.301]		[0.424]		[0.407]	
T1 (machine learning)	0.009*	0.242*	0.005	0.128	0.010*	0.269*	-0.009*	-0.233*
	(0.005)	(0.130)	(0.004)	(0.105)	(0.005)	(0.149)	(0.005)	(0.138)
T2 (business as usual)	-0.006	-0.254	-0.004	-0.143	-0.010**	-0.390**	0.008	0.338
	(0.004)	(0.173)	(0.004)	(0.147)	(0.005)	(0.191)	(0.005)	(0.220)
Pr($\beta[T1] = \beta[T2]$=0)	0.066	0.057	0.305	0.284	0.037	0.032	0.074	0.066
Pr($\beta[T1] = \beta[T2]$)	0.021	0.021	0.127	0.127	0.011	0.009	0.024	0.025

	Index (s.d.)		College quality (s.d.)		Index drop non-admitted (s.d.)		Undermatch (=1)	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(5)		(6)		(7)		(8)	
Control mean	-0.069		-0.170		-0.085		0.280	
Control sd	[0.975]		[1.202]		[0.980]		[0.449]	
T1 (machine learning)	0.022**	0.598**	0.028**	0.770**	0.010	0.262	-0.011*	-0.288*
	(0.009)	(0.255)	(0.013)	(0.360)	(0.007)	(0.173)	(0.006)	(0.166)
T2 (business as usual)	-0.007	-0.265	-0.009	-0.337	0.002	0.081	0.007	0.285
	(0.008)	(0.330)	(0.012)	(0.483)	(0.005)	(0.189)	(0.006)	(0.226)
Pr($\beta[T1] = \beta[T2]$=0)	0.042	0.035	0.077	0.066	0.311	0.303	0.056	0.048
Pr($\beta[T1] = \beta[T2]$)	0.018	0.027	0.035	0.053	0.336	0.466	0.017	0.019

Notes: This table reports the OLS regression results of the ITT effects (Model (2)) and TOT effects (Model (4)) of the individualized advising interventions in 2017 on a family of college access and match outcomes. The sample includes the universe of public high school graduates in Ningxia in 2017 (N=54,055). Column (7) only includes students who were admitted to college (N=45,482). **Machine learning** intervention used the predicted probabilities of admissions to each college-major-order for each student and other automatic data analyses to assist the conventional online personalized advising. **Business as usual** intervention provided low-touch, brief college application guidelines and tips to students, which is used as a placebo test that mimics a low-price for-profit consulting service available to students if they are willing to pay for it. **Index** denotes a principal-component index of college quality using information from five measures (median, mean, and minimum admissions scores; national college ranking scores and percentiles). **College quality** is the national college ranking score (standardized) using college admissions data from 1996-2017 and administrative data on institutional resources for every college in China. The other outcomes are dichotomous variables. All regressions control for student-level covariates and strata fixed effects. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 4: ITT and TOT effects of individualized advising programs: College application behaviors

Panel A: Predicted outcomes	*Ex-ante* applied college quality		*Ex-ante* mean admissions prob.		*Ex-ante* expected college quality		*Ex-post* applied college quality	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1)		(2)		(4)		(4)	
Control mean	-0.432		0.646		-0.417		-0.219	
Control sd	[1.329]		[0.138]		[1.031]		[1.133]	
T1 (machine learning)	0.025**	0.683**	-0.003*	-0.078*	0.023**	0.638**	0.020**	0.549**
	(0.011)	(0.309)	(0.002)	(0.044)	(0.011)	(0.291)	(0.009)	(0.259)
T2 (business as usual)	-0.008	-0.328	0.001	0.048	-0.008	-0.314	-0.009	-0.353
	(0.013)	(0.522)	(0.002)	(0.075)	(0.012)	(0.464)	(0.008)	(0.339)
Pr(T1=T2=0)	0.064	0.055	0.181	0.174	0.068	0.057	0.046	0.040
Pr(T1=T2)	0.047	0.074	0.112	0.144	0.042	0.064	0.016	0.021
Panel B: Strategies	Apply to colleges in ML list		College application Index		Strategy Targeting		Strategy General nudge	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(5)		(6)		(7)		(8)	
Control mean	0.305		-0.030		-0.028		-0.029	
Control sd	[0.460]		[1.001]		[0.994]		[1.003]	
T1 (machine learning)	0.011**	0.306**	0.022*	0.599*	0.033***	0.894***	0.030**	0.804**
	(0.006)	(0.150)	(0.012)	(0.332)	(0.011)	(0.305)	(0.012)	(0.338)
T2 (business as usual)	-0.003	-0.114	-0.018	-0.728*	-0.026**	-1.063**	-0.012	-0.489
	(0.006)	(0.225)	(0.011)	(0.439)	(0.013)	(0.491)	(0.012)	(0.482)
Pr(T1=T2=0)	0.136	0.119	0.126	0.101	0.007	0.005	0.058	0.049
Pr(T1=T2)	0.099	0.139	0.043	0.035	0.002	0.002	0.035	0.042
Panel C: Preferences	Special programs		Tuition & quota		Location		Major	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(9)		(10)		(11)		(12)	
Control mean	-0.030		0.030		-0.058		0.003	
Control sd	[0.978]		[1.012]		[0.977]		[0.999]	
T1 (machine learning)	-0.005	-0.137	0.018	0.484	-0.005	-0.137	-0.002	-0.058
	(0.017)	(0.456)	(0.014)	(0.375)	(0.011)	(0.294)	(0.016)	(0.425)
T2 (business as usual)	-0.000	-0.010	-0.001	-0.014	-0.003	-0.113	-0.009	-0.354
	(0.012)	(0.499)	(0.011)	(0.448)	(0.010)	(0.395)	(0.017)	(0.672)
Pr(T1=T2=0)	0.956	0.955	0.464	0.434	0.878	0.874	0.864	0.861
Pr(T1=T2)	0.801	0.831	0.322	0.390	0.879	0.958	0.774	0.714

Notes: This table reports the OLS regression results of the ITT effects (Model (2)) and TOT effects (Model (4)) of the individualized advising interventions in 2017 on a family of college application behaviors. The sample includes the universe of public high school graduates in Ningxia in 2017 (N=54,055). **Machine learning** intervention used the predicted probabilities of admissions to each college-major-order for each student and other automatic data analyses to assist the conventional online personalized advising. **Business as usual** intervention provided low-touch, brief college application guidelines and tips to students, which is used as a placebo test that mimics a low-price for-profit consulting service available to students if they are willing to pay for it. See the text for more descriptions of the outcome variables. All regressions control for student-level covariates and strata fixed effects. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 5: Effects of machine learning elements on improving college choices and admissions

Panel A: College choice behaviors	(1) Mean prob. of applied colleges	(2) Number of "out of reach" colleges	(3) Number of reach colleges	(4) Number of match colleges	(5) Number of safety colleges
Control mean	0.254	4.386	0.095	0.188	0.244
	[0.320]	[2.006]	[0.332]	[0.516]	[0.702]
T1 (Human expertise)	0.033	-0.210	0.018	0.041	-0.019
	(0.024)	(0.147)	(0.026)	(0.041)	(0.049)
T2 (Data)	0.110***	-0.694***	0.008	0.084*	0.170***
	(0.024)	(0.149)	(0.027)	(0.045)	(0.061)
T3 (Human expertise+Data)	0.185***	-1.191***	0.041	0.218***	0.223***
	(0.023)	(0.148)	(0.027)	(0.053)	(0.057)
T4 (Human expertise+Data+Classification)	0.143***	-1.010***	0.137***	0.217***	0.109**
	(0.023)	(0.148)	(0.033)	(0.052)	(0.053)
Pr($\beta[T1] = \beta[T4]$)	0.000	0.000	0.000	0.001	0.013
Pr($\beta[T2] = \beta[T4]$)	0.177	0.039	0.000	0.018	0.327
Pr($\beta[T3] = \beta[T4]$)	0.084	0.234	0.006	0.977	0.058

Panel B: College admissions results	(6) Admitted	(7) Possible admissions offers	(8) Ranking of admitted college	(9) 2020 admission score of admitted college (sd)	(10) Individual ranking in the AI competition
Control mean	0.451	1.448	424.295	0.661	2,001.301
	[0.498]	[1.926]	[165.492]	[0.738]	[1,103.758]
T1 (Human expertise)	0.044	0.151	-11.140	0.060	-69.835
	(0.036)	(0.142)	(11.871)	(0.053)	(78.688)
T2 (Data)	0.106***	0.602***	-36.952***	0.158***	-228.987***
	(0.035)	(0.144)	(11.912)	(0.052)	(77.385)
T3 (Human expertise+Data)	0.211***	1.027***	-65.964***	0.320***	-443.644***
	(0.034)	(0.142)	(11.704)	(0.050)	(75.409)
T4 (Human expertise+Data+Classification)	0.174***	0.759***	-59.007***	0.269***	-374.858***
	(0.034)	(0.141)	(12.057)	(0.051)	(77.509)
Pr($\beta[T1] = \beta[T4]$)	0.000	0.000	0.000	0.000	0.000
Pr($\beta[T2] = \beta[T4]$)	0.042	0.285	0.066	0.028	0.055
Pr($\beta[T3] = \beta[T4]$)	0.264	0.066	0.555	0.300	0.354

Notes: This table reports the effects of accessing different information (classified as machine learning elements) on college choice behaviors and simulated admissions outcomes in a survey experiment. The sample included 2,542 Chinese high school graduates in 2020. **Mean probability of applied colleges** in column 1 averages the admissions probabilities (generated by the causal forest model used for this subsection) of all the six colleges that each participant applied to. The types of colleges in columns 2-4 are based on predicted admissions probabilities. **Admitted** in column 6 simulates admissions results using the same rules in the contextual province that assigned at most one offer to each student. **Possible admissions offers** in column 7 simulates admissions results by assuming multiple offers are possible. **College ranking** in column 8 uses the same ranking information as in Table 3. **Individual ranking in the AI competition** ranks each participant based on the simulated admissions results from 1 to 2542. All regressions control for gender, parental education, STEM track, risk measure, and an indicator of understanding the college admissions mechanism. Robust standard errors are in parentheses. * significant at 10%, ** significant at 5%, *** significant at 1%.

A Appendix figures

Figure A.1: Location of Ningxia

Notes: Ningxia, officially the Ningxia Hui Autonomous Region, has the third smallest GDP in China with Muslims forming more than 38% of its population. Most of the region is desert, making Ningxia one of the poorest provinces in northwestern China. In 2017, the annual per capita disposable (after tax) income of urban residents is about $4,200 (national average: $5,600), and that of rural residents is $1,650 (national average: $2,060). About 800,000 of its 6 million population are under the poverty line that earn less than $1 a day.

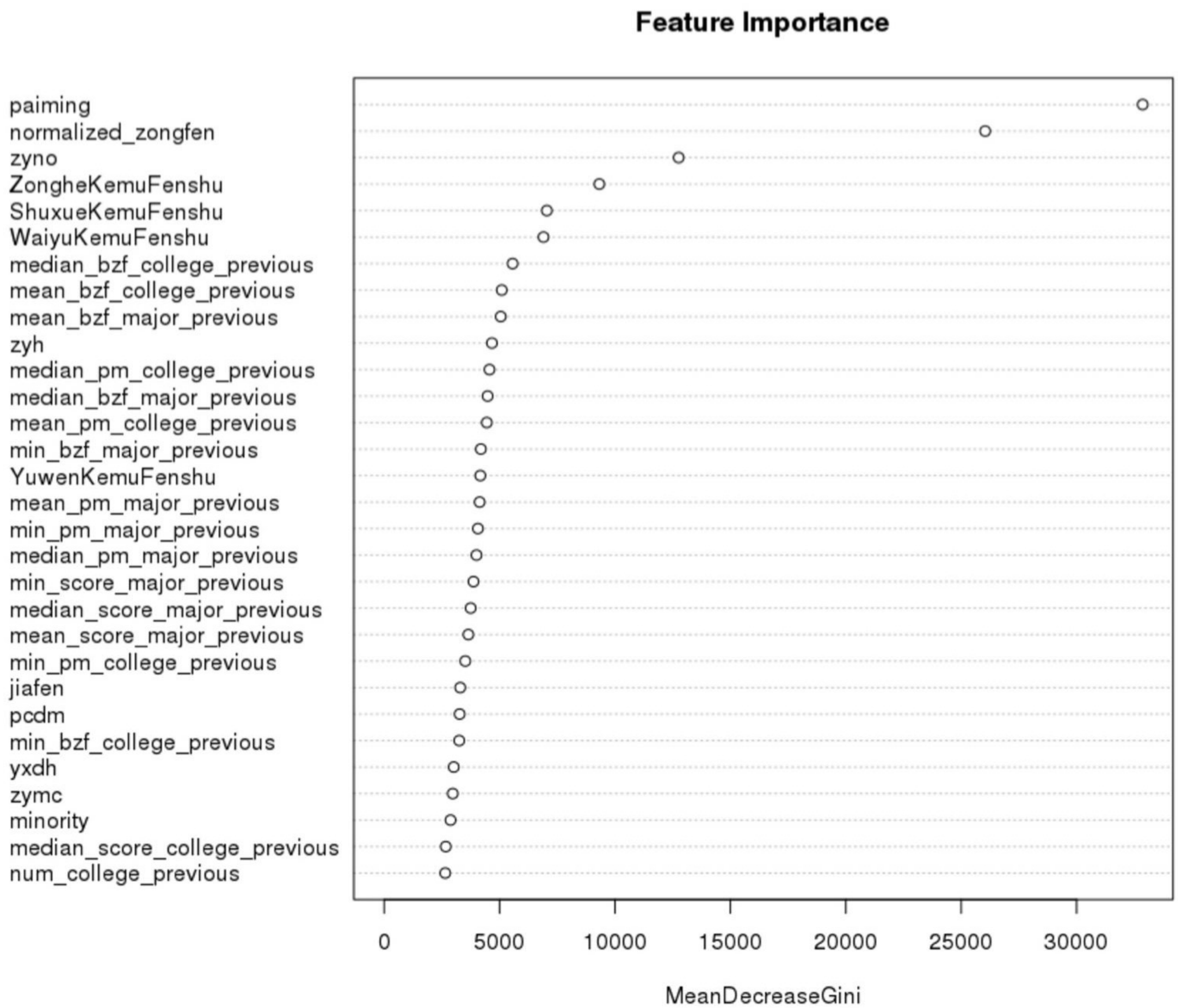


Figure A.2: Relative importance of features in the random forest predictions

Notes: This figure plots feature importance of the random forest model from the original output graph. We used student-college-major-rank order level data in 2015 to train the model (80% training set and 20% test set). The prediction accuracy was 94.3%. The most important two features are within province-track CEE score ranking (*paiming*) and within province-track CEE score (*normalized_zongfen*). Students with the same total CEE score may have different ranking because the differences in their subject scores (ranking weight order: track composite, Chinese, math, English for non-STEM students; track composite, math, Chinese, English for STEM students). The third important feature is the college rank order list (*zyno*).

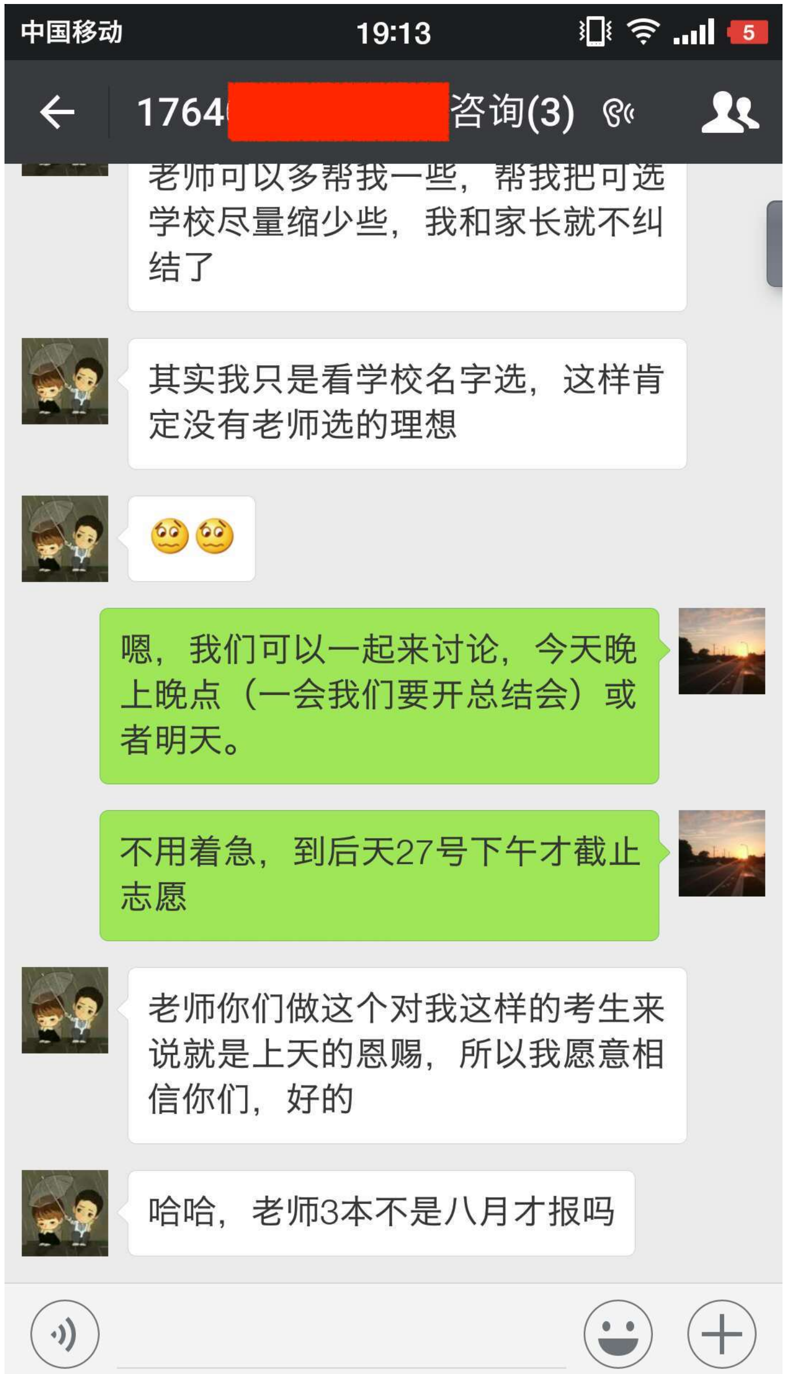


Figure A.3: Example of the online individualized advising in 2017

Notes: The conversations show two facts: This student would simply choose college by names (a behavioral mistake), and he was in need of a short list to assist his college applications.
Translation of the conversations:
Student: Teacher, you could provide me a short list of colleges that my parents and I will not be entangled with the choices of colleges.
Student: Honestly, I would just choose colleges by their names. Your choices must be better than mine.
Student: (Smile)
Advisor: Yes, we can discuss about your applications together, later tonight or tomorrow (will have a meeting right now).
Advisor: Don't worry, June 27 is the deadline (the day after tomorrow).
Student: Teacher, your assistance is god's grace for students like me. I trust you.

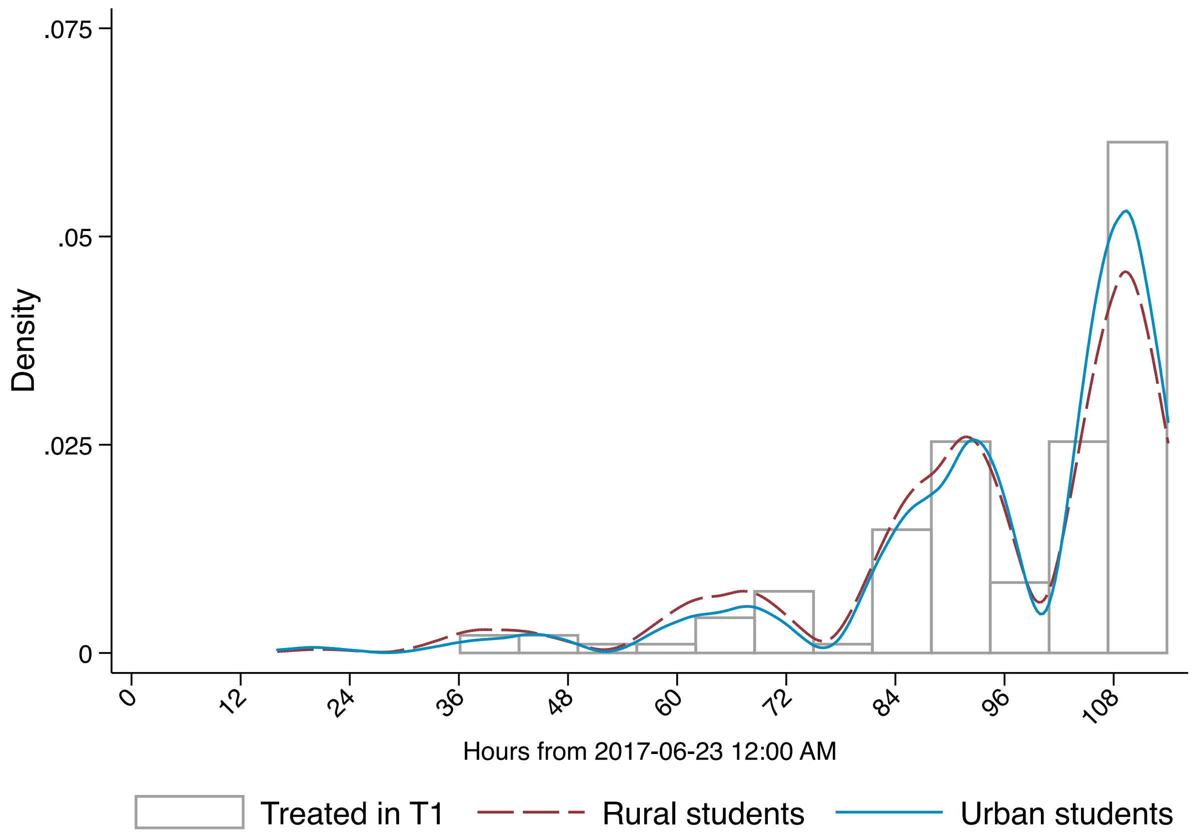


Figure A.4: Distribution of college application submission time

Notes: This figure shows the kernel distribution of college application submission time, separately for rural and urban students in the control group who were eligible for applying to selective colleges. The gray bar shows the distribution of submission time of students who were assigned to receive the "machine learning" advising and eventually received the advising. College application was open from 2017-06-23 16pm to 2017-06-27 18pm.

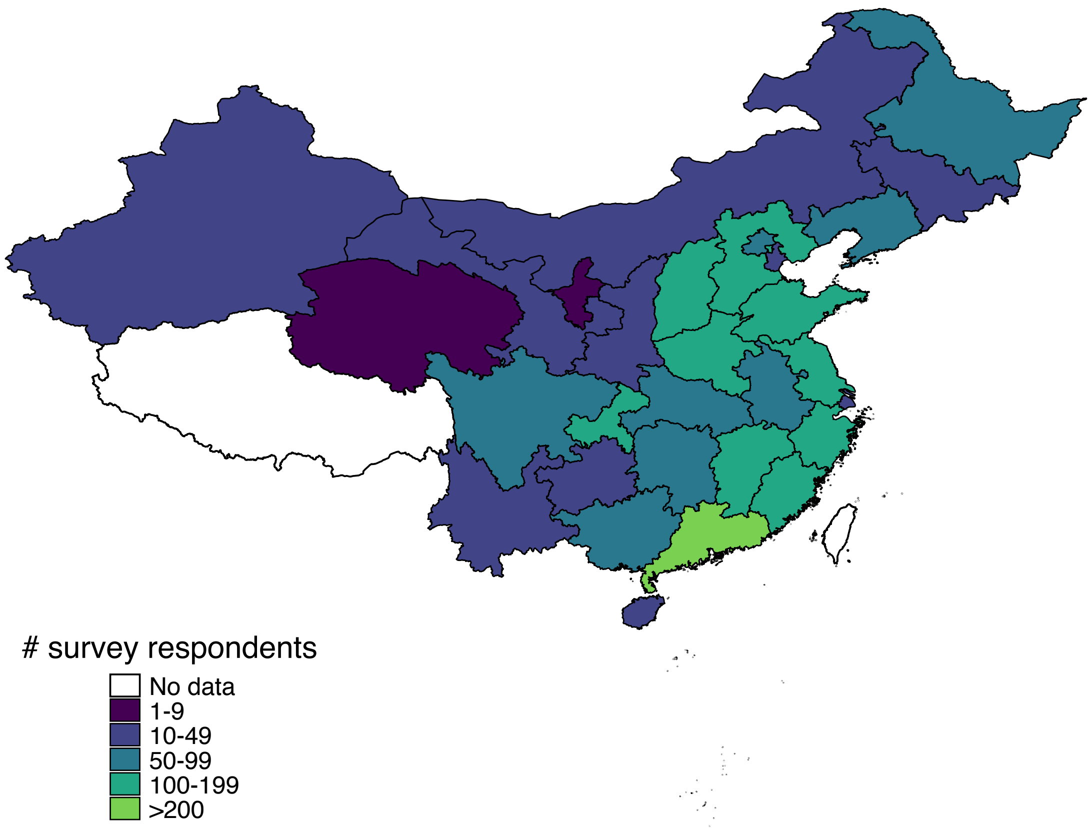


Figure A.5: Geographic distribution of the survey respondents in 2020

Notes: This figure plots geographic distribution of the survey respondents (high school graduates) in 2020.

(a) Cover page with the competition rules

(b) Control group

(c) Treatment 1: Human expertise

(d) Treatment 2: Data

(e) Treatment 3: Human expertise + Data

(f) Treatment 4: Human expertise + Data + Classification

Figure A.6: Survey experimental design in 2020

Notes: This figure shows the survey experiment in 2020. The left part of each subfigure shows the the screen shots from the survey website. The left part of each subfigure highlight the key information. The title line of each subfigure is "Competing with AI to win an iPad: Looking for college application experts."

B Appendix tables

Table B.1: The poverty gap in college match

	Outcome: Index of college match			
	(1)	(2)	(3)	(4)
Rural	-0.134***	-0.149***	-0.107***	-0.063***
	(0.015)	(0.014)	(0.011)	(0.007)
Female		-0.049***	-0.050***	-0.044***
		(0.010)	(0.009)	(0.008)
Minority		-0.008	-0.081***	-0.061***
		(0.016)	(0.011)	(0.010)
Age		0.003	-0.005	-0.003
		(0.009)	(0.009)	(0.009)
STEM		0.186***	0.199***	0.194***
		(0.013)	(0.011)	(0.010)
Repeater		0.178***	0.068***	0.095***
		(0.017)	(0.013)	(0.014)
CEE score	0.811***	0.809***	0.786***	0.764***
	(0.017)	(0.019)	(0.013)	(0.015)
School FE	No	No	No	Yes
Observations	39,385	39,385	35,332	35,332
R-squared	0.630	0.646	0.713	0.719

Notes: This table reports the OLS regression (Model (1)) results of the rural-urban gap in college match outcomes (as being summarized in the single index), using the control group sample in 2017. Columns (3) and (4) exclude students who were not admitted to a college. Columns (4) controls for high school fixed effects. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table B.2: Balance checks

	All students			Excluding students in treated teachers' classes		
		Diff from Control			Diff from Control	
	Control (1)	T1 (2)	T2 (3)	Control (4)	T1 (5)	T2 (6)
Rural	0.572	0.000	0.000	0.570	0.000	0.000
	[0.495]	(0.000)	-	[0.495]	(0.000)	(0.000)
Female	0.549	0.000	0.000	0.549	0.000	0.000
	[0.498]	(0.000)	(0.000)	[0.498]	(0.000)	(0.000)
Minority	0.321	0.002	-0.001	0.317	0.002	-0.002
	[0.467]	(0.002)	(0.001)	[0.465]	(0.002)	(0.001)
Age	0.873	-0.010	-0.001	0.874	-0.010	-0.002
	[0.333]	(0.006)	(0.004)	[0.332]	(0.007)	(0.004)
STEM	0.665	0.001	-0.001	0.661	0.002	-0.003
	[0.472]	(0.004)	(0.004)	[0.473]	(0.004)	(0.004)
Repeater	0.231	0.008	0.002	0.236	0.007	0.003
	[0.422]	(0.005)	(0.005)	[0.425]	(0.005)	(0.005)
CEE score	0.034	0.000	-0.001	0.040	0.007	0.009
	[0.943]	(0.011)	(0.011)	[0.944]	(0.011)	(0.011)
F test		1.124	0.091		0.770	0.443
(P value)		0.360	0.999		0.615	0.871
Students	43,038	5,647	5,370	39,385	5,246	4,935
Schools	61	61	61	61	61	61

Notes: This table reports the balance checks results using student-level data in 2017. There were 836 treated students in treated teachers' classes. **Age** is a dummy indicator for students younger than nineteen years old by the time of the college entrance exam. **Repeater** is a dummy indicator for students having taken the CEE at least once in the previous years. **CEE score** is standardized by STEM/non-STEM tracks using the full sample. Joint F test results are from regressions in Table B.3. Strata fixed effects are included. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table B.3: Balance checks: Prediction of treatment status

	All students		Excluding students in treated teachers' classes	
	T1	T2	T1	T2
	(1)	(2)	(1)	(2)
Rural	0.010	-0.003	0.007	-0.011
	(0.009)	(0.009)	(0.009)	(0.008)
Female	-0.003	-0.000	-0.003	0.001
	(0.003)	(0.002)	(0.003)	(0.003)
Minority	0.021	-0.006	0.016	-0.023
	(0.018)	(0.019)	(0.019)	(0.016)
Age	-0.009	-0.002	-0.009	-0.002
	(0.006)	(0.004)	(0.006)	(0.004)
STEM	0.001	-0.002	0.003	-0.004
	(0.006)	(0.006)	(0.006)	(0.006)
Repeater	0.007*	0.002	0.005	0.001
	(0.004)	(0.004)	(0.004)	(0.004)
CEE score	-0.001	-0.001	0.000	0.001
	(0.002)	(0.002)	(0.002)	(0.002)
F test	1.124	0.091	0.770	0.443
(P value)	0.360	0.999	0.615	0.871
Observations	48,685	48,408	44,631	44,320
R-squared	0.121	0.019	0.122	0.020

Notes: This table reports the balance checks results from separate OLS regressions that predict the treatment status using student-level data in 2017. Strata fixed effects are included. Joint F test results are reported. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table B.4: ITT and TOT effects of individualized advising programs: Itemized outcomes in the index measure

	College median (s.d.)		College mean (s.d.)		College min (s.d.)		Ranking (pctl)	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1)		(2)		(3)		(4)	
Control mean	-0.227		-0.200		-0.998		45.121	
Control sd	[1.159]		[1.109]		[1.317]		[33.882]	
T1 (machine learning)	0.021**	0.565**	0.019**	0.511**	0.032**	0.874**	0.833**	22.703**
	(0.010)	(0.284)	(0.009)	(0.245)	(0.013)	(0.369)	(0.378)	(10.425)
T2 (business as usual)	-0.008	-0.314	-0.008	-0.306	-0.003	-0.099	-0.314	-12.526
	(0.009)	(0.375)	(0.008)	(0.315)	(0.012)	(0.484)	(0.340)	(13.743)
Pr($\beta[T1] = \beta[T2]$=0)	0.071	0.063	0.052	0.046	0.060	0.051	0.055	0.048
Pr($\beta[T1] = \beta[T2]$)	0.027	0.038	0.018	0.025	0.047	0.082	0.022	0.031

Notes: This table reports the OLS regression results of the ITT effects (Model (2)) and TOT effects (Model (4)) of the individualized advising interventions in 2017 on itemized college match outcomes (except for college quality in column 6 of Table 3) that build the single index. The sample includes the universe of public high school graduates in Ningxia in 2017 (N=54,055). **Machine learning** intervention used the predicted probabilities of admissions to each college-major-order for each student and other automatic data analyses to assist the conventional online personalized advising. **Business as usual** intervention provided low-touch, brief college application guidelines and tips to students, which is used as a placebo test that mimics a low-price for-profit consulting service available to students if they are willing to pay for it. All regressions control for student-level covariates and strata fixed effects. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table B.5: ITT and TOT effects of individualized advising programs: Itemized college application behaviors

College application behaviors	Control (1)	Effects of T1 ITT (2)	Effects of T1 TOT (2)	Effects of T2 ITT (3)	Effects of T2 TOT (3)
		A. Strategy - Targeting			
Apply to at least one college in match tier	0.757	0.008*	0.221*	-0.008	-0.320
	[0.429]	(0.005)	(0.129)	(0.005)	(0.216)
Estimated gap within 0.15 s.d.	0.282	0.013**	0.348**	0.002	0.074
	[0.450]	(0.005)	(0.152)	(0.006)	(0.246)
Descending order list	0.194	0.013**	0.355**	-0.003	-0.109
	[0.395]	(0.006)	(0.163)	(0.005)	(0.205)
First listed college is “reach"	0.617	0.016***	0.421**	-0.014**	-0.581**
	[0.486]	(0.006)	(0.165)	(0.007)	(0.274)
Last listed college is “safety”	0.264	0.004	0.106	-0.003	-0.129
	[0.441]	(0.006)	(0.162)	(0.006)	(0.258)
Combined “reach,” “match,” “safety”	0.216	0.011*	0.286*	-0.014**	-0.577**
	[0.411]	(0.006)	(0.171)	(0.007)	(0.276)
Percent of “reach” colleges	28.459	0.527	14.693	0.093	4.046
	[30.401]	(0.409)	(11.239)	(0.371)	(15.131)
Percent of “match” colleges	36.465	0.165	3.900	-0.587	-24.020*
	[30.266]	(0.418)	(11.521)	(0.357)	(14.351)
Percent of “safety” colleges	35.076	-0.692	-18.593	0.494	19.974
	[33.510]	(0.476)	(13.288)	(0.456)	(17.900)
		B. Strategy - General nudge			
Apply to all four colleges in match tier	0.669	0.011*	0.286*	-0.005	-0.216
	[0.470]	(0.006)	(0.155)	(0.006)	(0.244)
Percent of majors applied to	61.181	0.740**	19.911**	-0.498	-20.134
	[30.411]	(0.362)	(9.989)	(0.363)	(14.702)
Percent of flexible major assignment	61.475	1.001*	27.646*	-0.056	-1.892
	[41.815]	(0.592)	(16.318)	(0.459)	(18.517)

Notes: This table reports the OLS regression results of the ITT effects (Model (2)) and TOT effects (Model (4)) of the individualized advising interventions in 2017 on itemized college application behaviors. Results for the rest strategies and preferences are statistically insignificant. The sample includes the universe of public high school graduates in Ningxia in 2017 (N=54,055). **Machine learning** intervention used the predicted probabilities of admissions to each college-major-order for each student and other automatic data analyses to assist the conventional online personalized advising. **Business as usual** intervention provided low-touch, brief college application guidelines and tips to students, which is used as a placebo test that mimics a low-price for-profit consulting service available to students if they are willing to pay for it. See the text for more descriptions of the outcome variables. All regressions control for student-level covariates and strata fixed effects. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table B.6: Correlations between application time use and outcomes in college choices and admissions

	Admission (=1)				Match index (s.d.)			
	OLS	OLS School FE	IPW+RA	IV	OLS	OLS School FE	IPW+RA	IV
		(1)				(2)		
Hours	0.005***	0.005***		0.009**	0.016***	0.008***		0.017**
	(0.000)	(0.000)		(0.004)	(0.001)	(0.000)		(0.007)
Later than 2 days	0.249***	0.225***	0.213***	0.704**	0.890***	0.363***	0.351***	1.351**
	(0.010)	(0.009)	(0.003)	(0.360)	(0.056)	(0.014)	(0.006)	(0.648)

	Application index (s.d.)				Targeting stragey (s.d.)			
	OLS	OLS School FE	IPW+RA	IV	OLS	OLS School FE	IPW+RA	IV
		(3)				(4)		
Hours	0.014***	0.013***		0.023**	0.005***	0.003***		0.034***
	(0.001)	(0.001)		(0.010)	(0.001)	(0.000)		(0.012)
Later than 2 days	0.781***	0.629***	0.503***	1.852**	0.708***	0.570***	0.576***	2.740***
	(0.038)	(0.028)	(0.007)	(0.860)	(0.036)	(0.023)	(0.010)	(1.034)

Notes: This table reports the correlation between application time use the outcomes in college choices and admissions, using four different strategies: OLS without school fixed effects; OLS with school fixed effects; inverse-probability-weighted regression adjustment; and IV (using the random assignment to the two advising interventions as instrumental variables). Each cell is from a separate regression. All regressions control for student-level covariates. **IPW+RA** controls for high school fixed effects. **IV** includes strata fixed effects. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table B.7: Heterogeneity in the ITT effects on time spent for application (hours)

	All	Urban (1)	Rural (2)	Male (3)	Female (4)
N (students)	54,055	23,776	30,279	24,658	29,397
Control	57.26	60.516	54.819	57.282	57.238
	[36.916]	[37.196]	[36.516]	[37.280]	[36.615]
T1 (machine learning)	0.973**	0.695	1.249*	1.588**	0.396
	(0.431)	(0.554)	(0.648)	(0.706)	(0.654)
T2 (business as usual)	-0.777*	-0.649	-0.892	-0.893	-0.676
	(0.458)	(0.743)	(0.669)	(0.642)	(0.588)

	Non-minority (5)	Minority (6)	Low-achieving (7)	High-achieving (8)
N (students)	36,773	17,282	40,805	13,250
Control	56.985	57.835	48.381	90.014
	[37.243]	[36.209]	[34.280]	[26.299]
T1 (machine learning)	0.916*	1.090	1.393**	0.440
	(0.556)	(0.690)	(0.595)	(0.514)
T2 (business as usual)	-1.450***	0.663	-0.732	-0.520
	(0.523)	(0.756)	(0.576)	(0.612)

Notes: This table reports the ITT effects of individualized advising interventions on application time use. **The outcome variable** is the total hours from the open dates (June 23 for selective colleges and August 1 for non-selective colleges). Students who did not submit their applications were coded as zero hours (results are similar if excluding these students). **Machine learning** intervention used the predicted probabilities of admissions to each college-major-order for each student and other automatic data analyses to assist the conventional online personalized advising. **Business as usual** intervention provided low-touch, brief college application guidelines and tips to students, which is used as a placebo test that mimics a low-price for-profit consulting service available to students if they are willing to pay for it. All regressions control for student-level covariates and strata fixed effects. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table B.8: Heterogeneous ITT effects of individualized advising programs

	Urban	Rural	Male	Female	Non-minority	Minority	Low achieving	High achieving
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Take-up	0.039***	0.033***	0.033***	0.039***	0.037***	0.035***	0.027***	0.048***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.003)	(0.004)
Admission	0.008	0.010	0.022***	-0.003	0.008	0.011	0.015*	0.004***
	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)	(0.008)	(0.001)
College match index	0.012	0.031**	0.040***	0.005	0.025**	0.014	0.032**	0.014*
	(0.012)	(0.012)	(0.012)	(0.013)	(0.012)	(0.013)	(0.013)	(0.007)
ML list	0.007	0.015*	0.022***	0.001	0.014**	0.004	0.014**	0.004
	(0.009)	(0.009)	(0.007)	(0.008)	(0.007)	(0.010)	(0.007)	(0.007)
College choice index	0.029*	0.016	0.046**	0.001	0.032**	0.002	0.041*	0.002
	(0.015)	(0.018)	(0.018)	(0.019)	(0.016)	(0.018)	(0.021)	(0.010)
Targeting strategy	0.036**	0.031*	0.058***	0.011	0.046***	0.006	0.054**	-0.001
	(0.016)	(0.019)	(0.019)	(0.018)	(0.015)	(0.021)	(0..020)	(0.014)
N (students)	23,776	30,279	24,658	29,397	36,773	17,282	40,805	13,250

Notes: This table reports heterogeneous ITT effects of the **machine learning** assisted advising on college choice and admission outcomes. The left column lists the outcomes (except for the number of observations in the first row). The sample includes the universe of public high school graduates in Ningxia in 2017 (N=54,055). **Machine learning** treatment applies the predicted probabilities of admissions to each college-major-order for each student and other automatic data analyses to assist the conventional online personalized advising. All regressions control for student-level covariates and strata fixed effects. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table B.9: Balance checks for the 2020 survey experiment

	(1) Control	(2) T1	(3) T2	(4) T3	(5) T4	ANOVA
Female	0.335	0.334	0.327	0.345	0.336	0.999
	[0.473]	[0.472]	[0.470]	[0.476]	[0.473]	
STEM track	0.552	0.548	0.574	0.579	0.617	0.985
	[0.498]	[0.498]	[0.495]	[0.494]	[0.487]	
Parent: High school	0.374	0.366	0.400	0.404	0.385	0.993
	[0.484]	[0.482]	[0.490]	[0.491]	[0.487]	
Parent: College	0.487	0.493	0.456	0.474	0.482	1.000
	[0.500]	[0.500]	[0.499]	[0.500]	[0.500]	
High risk (>5)	0.675	0.687	0.664	0.672	0.668	0.995
	[0.469]	[0.464]	[0.473]	[0.470]	[0.471]	
Understand rule	0.618	0.626	0.613	0.615	0.594	0.998
	[0.486]	[0.484]	[0.487]	[0.487]	[0.492]	
Observations	495	473	535	527	512	$\{0.974\}^{\#}$

Notes: This table reports the balance test results for the 2020 survey experiment. Columns 1-5 reports mean and standard deviation for each covariate in each randomized group. The last columns reports p-values from ANOVA for each covariate across groups. # is the p-value from a multinomial logit model that regresses the group indicators on all the covariates. **High risk** indicates that participants' self-reported risk-taking score is above 5 in a 1-10 scale. **Understand rule** indicates that participants correctly answered a survey question about the Deferred Acceptance mechanism.

C Additional descriptions

C.1 Measuring college application behaviors and their correlations with college admissions

C.1.1 Tier-specific college-major applications in Chinese centralized admissions

College-major applications and admissions in China proceed by institutional selectivity tiers within province-track.[40] Each college-major belongs to a predetermined tier (a college may have majors in different tiers). A student's eligibility to apply to colleges in each tier is mostly determined by her CEE score. She could apply to Tier 1 if and only if her CEE score is above the tier-specific cutoff score. She can also apply to the other tiers. A student could only apply to Tier 4 colleges if her CEE score is below Tier 3 cutoff. Few students could not apply to any college with CEE score below the very low Tier 4 cutoff (200 raw points out of 750).

Table C.1 shows a simplified version of the college application form in Ningxia in 2016. On the one hand, the application (administrative) process is simplified. Many common requirements in decentralized admissions systems (e.g., score-sending, institution-specific essays, AP courses, reference letters) are no longer needed. Students need to choose colleges and majors of their interests from the pull-down menu in the online application system. If they already have a list of interested and majors at hand, they can finish the application process in minutes. On the other hand, the application is complicated. Students would have to consider every cell in the application form in Table C.1. They need to build knowledge and skills to pick colleges and majors strategically. Therefore, a knowledge-based intervention on the use of college choice knowledge and skills would improve students' applications and admissions.

The application form corresponds to the order of admissions. Within each institutional tier, there are several special programs that could be seen as sub-tiers within each tier. For instance, in addition to the primary Tier 1 (choice of four colleges), students who are eligible for Tier 1 admissions could potentially apply to (1) Tier 1 - Early Admissions, (2) Tier 1 - National Affirmative Action Programs for Rural Poor Students, (3) Tier 1 - Provincial Affirmative Action Programs for

[40] *Note:* This subsection is adopted from Ye (2020).

Rural Poor Students, (4) Tier 1 - Affirmative Action Programs for Minority Students, and (5) Tier 1 - Other Special Programs (e.g., College-level Affirmative Action Programs for Rural Poor Students). In Ningxia in 2016, a student, in theory, could apply to 58 different colleges (out of about 1,200 colleges) and then 348 college-major options (out of about 20,000).[41]

[41] There are 2,631 colleges in China (not including military colleges; till May 2017). But not all of them admit students from Ningxia.

Table C.1: College-major application form in Ningxia in 2017 (Simplified)

ID: Name: Track:

Tier	No.	College	Major						Flexible assignment?
			1	2	3	4	5	6	
Tier 1 - Early admissions	1								
	2								
Tier 2 - EA	1								
	2								
Tier 1 - National Affirmative Action (Rural)	A								
	B								
	C								
Tier 1	A								
	B								
	C								
	D								
Tier 1 - Provincial AA (Rural)	A								
	B								
Tier 1 - AA (Minority)	A								
	B								
	C								
Tier 1 - Special majors	1								
Tier 2	A								
	B								
	C								
	D								
Tier 2 - AA (Minority)	A								
	B								
	C								
Tier 2 - Special majors	1								
Tier 3	A								
	B								
	C								
	D								
Tier 3 - AA (Minority)	A								
	B								
	C								
Tier 4 - EA	1								
Tier 4	A								
	B								
	C								
	D								

Notes: This table adopts the original Chinese version of the application form and excludes a few rows of special program lists. In Ningxia in 2016, a student, in theory, could apply to 58 different colleges and then 348 college-major options. Data source: Baidu Wenku. Numbers in the "No." column indicates the admissions are based on the Boston Mechanism, and letters in that column indicates the admissions are based on the DA (Parallel) Mechanism.

C.1.2 Measuring college choice behaviors using students' application "big data"

Based on features of the tier-specific applications in the Chinese centralized college admission system, we focus on two sets of strategies. These strategies are expected to capture some of the key application behaviors for a knowledgeable and skillful student. We have also covered these strategies in our interventions of the application guidebook, school workshop, and personalized advising.

The first set of variables describe the targeting strategies that students should use to apply to a reasonable combination of peer, reach/match and safety colleges (and majors). These strategies require intensive knowledge and sophistication to make the accurate predictions of college-major admissions probabilities and apply for college-majors based on the predictions. This set of strategies are the key elements of our behavioral interventions as well as the data analysis in a students' college choice and application.

Many students do not understand the underlying mechanisms of college admissions that only rank (but not raw score) matters. They naively compare their CEE score in this year with college admissions raw scores, which results in large errors of identifying college types. Students may use different strategies in different tiers, but we use their behaviors in their match tier to represent their general knowledge and skills in college applications. A match tier is the highest possible institutional selectivity tier that a student qualifies for, which is similar to the use of selectivity tiers in defining undermatch in the literature (e.g., Smith, Pender and Howell, 2013). Besides, we focus on college-level application behaviors, but those choices of majors within each college is also worth exploring in the future research.

- **[Strategy 1.1] Estimated gap (within 0.15 s.d.).** <u>The behavioral rationale</u> is that students should equate their CEE score to admissions scores in the previous years (i.e., transforming the raw scores to standardized scores). For example, suppose that the raw CEE scores are 500 and 550 for a student ranked 10,000 in 2016 and 2015, a student in 2016 with CEE score of 500 should then look at colleges with admissions scores around 550 in 2015. If she applied to colleges with admissions scores around 500 in 2015, she would be very much likely to undermatch. The raw scores vary dramatically over the years. Suppose that the raw CEE scores are 600 and 550 for a student ranked 10,000 in 2016 and 2015, if a student with CEE score of 600 in 2016 applied to colleges with admissions scores around 600 in 2015, she would not be likely to be admitted by an undermatched college, but being rejected by all of her applied colleges. We construct this variable by estimating the gap (difference) between one's CEE score in 2016 and the equated median score (from 2015 to 2016) of the college she listed in the second college choice in the match tier.[42] This variable equals to 1 if the estimated gap is within 0.15 s.d.. Sample mean is 34%. The strategy is that students need to acquire the knowledge of score equating (and the principle of why score equating is needed) as well as data of the crosswalks between raw scores and rankings over the years. They need to do the score equating by themselves before choosing colleges and majors to apply for.[43]

- **[Strategy 1.2] Apply to colleges in the match tier.** The behavioral rationale is that students would have access to most of their peer/match colleges in the match tier. Students may have behavioral mistakes of not applying to the match tier but only to colleges in lower tiers, or they only applied to special programs but not to colleges in the primary sub-tier. We construct this variable by identifying students who did not apply to colleges in match tier. Sample mean is 23% that about 23 percent of students in 2016 (in the untreated sample) did not apply to colleges in match tier. This number does not include those who did not submit their college applications.[44]
- **[Strategy 1.3] Apply to colleges without admissions data in the prior year.** The number of colleges that admit students in one province may change over time. Each year there

[42]We choose the second choice order as that it is expected that a student should apply to a match college in here second or third choice (first choice as a reach college and last choice as a safety choice). Results are very stable if we use other choices or a summary statistic of these choices.

[43]Though correctly centered, a large proportion of students apply to colleges that they would be substantially undermtached or overmatched. It is very likely because they do not (understand and) do score equating. From our fieldwork observations, high school teachers also lack the knowledge about score equating.

[44]For students who prefer low tuitions and are only eligible for Tier 3 and 4 colleges, one rational choice is that they may not be interested in colleges in Tier 3 (private four-year colleges with high tuitions) and only applied to Tier 4 colleges.

are “new” colleges for students to apply to. The behavioral rationale is that students need to infer/predict the admissions data in previous years for these “new” colleges using other information, and they may take risks of applying to these colleges. However, if most students are risk-averse and do not apply to those colleges, it is a good opportunity for skillful students to gain an overmatched admission. We construct this variable by identifying students who applied to colleges in the match tier without admissions data in the prior year. Sample mean is 2%.

- **[Strategy 1.4] Descending order of colleges in the match tier.** The behavioral rationale is that students should apply to a mix of reach, peer and safety colleges to maximize their opportunities of getting into reach and peer colleges, and to minimize the risks of being rejected by all (Hoxby and Avery, 2013). In order to correctly identify types of reach, peer and safety colleges, students need to understand the classification of these types (a rule of thumb is a 0.05-0.15 s.d. threshold) based on score-equating. Then, for the four college choices within each tier, given the institutional feature of Differed Acceptance (Parallel) mechanism, students should list their four choices in the descending order (choice A > choice B > choice C > choice D), otherwise any choices in higher orders with higher *ex post* admissions scores are meaningless. We construct this variable by a dichotomous indicator of students who did so in their match tier. Sample mean is 31%.
- **[Strategy 2.5] Targeting.** The behavioral rationale is that, although students are nudged to apply to a mix of reach, peer and safety colleges, they should not aim too high or too low. In other words, they need to have a tight range of colleges (centering around their CEE scores). We construct this variable by a dichotomous indicator of students with differences in college median score in the prior year between the first college choice and the last choice in the match tier in the range of (0, 0.5 s.d.). Sample mean is 35%.

The second set variables describe some general guidelines (or simple information/strategy):

- **[Strategy 2.1] Number of applied colleges.** The behavioral rationale is that increased applications are positively correlated with increased college opportunities (e.g., Pallais, 2015;

Hurwitz et al., 2017). However, applying to too many colleges without caution may result in undermatched colleges in some early admissions or special programs. A common mistake that we have observed in the field and from the data is that many Tier 1 eligible students incorrectly applied to colleges in "Tier 2 - Early Admissions." Colleges in "Tier 2 - Early Admissions admit students before those in "Tier 1" that these students missed their chances of much higher quality colleges in Tier 1. We construct this variable by counting the total number of all the colleges that a student applied to. Sample mean (using the untreated sample in 2016, see descriptions in the main text) is 7.2, with a minimum of 1 and a maximum of 40. The strategy is not deterministic that we recommend students to think about their applications carefully and the number of colleges to apply to is related to the targeting strategies in the second set variables.

- **[Strategy 2.2] Percent of applied majors.** The behavioral rationale is that, unless students are strongly against specific majors and they could bear the risks of being rejected by a college that considers her admission, students should fill in all the six major options within each college (or the maximum number of majors in that college). This is because that the college-then-major admissions give each student only one college temporary admission chance. If a student is eventually rejected by a college due to the unmatched of major applications, she will not be considered by other colleges in the same institutional tier and has to move down to lower tiers. In practice, many students only have strong major preferences, but do not understand the need for this strategy to reduce their rejection risks. We construct this variable by calculating the percent of major applications over total available major numbers given the colleges that a student applied to. Sample mean is 69.9%, with a minimum of 16.7% and a maximum of 100%.
- **[Strategy 2.3] Percent of flexible major assignment.** The behavioral rationale is that flexible major assignment minimizes the risks of being rejected by a college due to unmet major choices, which happens when all the majors within a college that a student applies to have higher admissions scores than her CEE score. If that student accepts flexible major

assignment within that college, then the college will assign her to a major that still has a spot (but that major may not be her interested one). The flexible assignment is actually to increase admission probability by sacrificing major preferences. We construct this variable by calculating the percent of college applications accepting flexible major assignment over the number of applied colleges. Sample mean is 69.2% with a minimum of 0 and a maximum of 100%. The strategy, which we strong nudged every student to use, is to accept a flexible major assignment at most of the applied colleges, if not all of them.

Student preferences and tastes are individual-specific and strictly unobservable. Particularly in constrained college applications, revealed preferences may not be precisely true. We construct four sets of proxy preferences using the application data.

The first set of preferences regard special programs that students may lack awareness and information and knowledge to understand these policies. One example is that the affirmative action programs for minority students vary greatly in college quality between national programs and in-province programs. Students may apply for both and end up with lower quality in-province colleges.

- **[Preference 1.1] Minority affirmative action programs.** The behavioral rationale is that students may lack information and knowledge to differentiate/understand different AA programs. National AA programs are of high quality (in selective colleges), but provincial AA programs are lower-quality. We construct this variable by identifying that if a student applied to any AA programs. Sample mean is 22%, with a minimum of 0 and a maximum of 1.
- **[Preference 1.2] Early admissions.** The behavioral rationale is that students may lack awareness of these programs and understanding of the policy. For example, the rural poor student affirmative action programs at selective colleges need pre-registry several months before CEE, but many students did not complete the registration. We construct this variable by identifying that if a student applied to any early admissions programs. Sample mean is 15%, with a minimum of 0 and a maximum of 1.

- **[Preference 1.3] Teachers' education.** The behavioral rationale is that these special teachers' education programs may be opportunities to enter higher quality colleges (based on one's CEE score). However, students may have strong major preferences. We construct this variable by counting the percent of applied majors in teacher's education. Sample mean is 5.2%, with a minimum of 1 and a maximum of 40.

The second set includes college tuition and quota, which are the primary information provided to students by the Department of Education.

- **[Preference 2] College tuition and quota.** The behavioral rationale is that low-income students may prefer low-tuition colleges, and risk-averse students may prefer college with larger admissions quota (Dynarski and Scott-Clayton, 2013; Hoxby and Avery, 2013; Loyalka, Wu and Ye, 2017). In China, selective colleges have lower tuitions than non-selective colleges. Within selectivity, tuitions vary across locations, college types and majors. Students may also use tuition as a naive indicator of college quality. College quota may be positively correlated with admissions probability (Kamada and Kojima, 2015), but students may be unaware of the quota information, which is provided to them by the Department of Education. We construct these variables by using median college tuition of all applied colleges and mean quota of all applied colleges. Sample mean of tuition is 6,300, with a minimum of 0 and a maximum of 40,700. Sample mean of quota is 708, with a minimum of 1 and a maximum of 2,993.

The third set of preference variables are the college location choices:

- **[Preference 3.1] Out-of-province colleges.** The behavioral rationale is that distance is one important factor shaping students' college choices, but focusing on in-province colleges would limit other high-quality college opportunities (Hillman, 2016; Hoxby, 2000; Long, 2004; Miller, 2017; Ovink et al., 2018). It is also true in Ningxia that high-quality colleges concentrate in the economically developed regions in China. Ningxia province, as a low-income region, lacks high-quality colleges. We construct this variable by calculating the percent of applied colleges locating in out-of-province regions (excluding economically advanced regions and Ningxia's neighborhood provinces). Sample mean is 38.8%, with a minimum of 0 and a maximum of 1.

- **[Preference 3.2] Out-of-province (advanced regions) colleges.**
  underlineWe construct this variable by calculating the percent of applied colleges locating in the most economically advanced regions of China, including Beijing, Shanghai, Guangdong. Sample mean is 6.6%, with a minimum of 0 and a maximum of 1.

The last set of preferences are major choices. We include the most popular ones (e.g., economics, computer science, international) and the least popular agricultural-related majors in the analytical variables.

- **[Preference 4] Majors.** We construct these variables by calculating the percent of each major group over the total number of applied majors. The mean values of Economics-related, Agricultural-related, Computer science-related, International-related, and Medical-related are 24.1%, 1.3%, 3.2%, 1.6%, 11.4%. We did not provide direct interventions on major choice but provided information about all the majors (e.g., coursework, college life, labor market outcomes). We nudged students to get to know each major well before making decisions. Additionally, this is also related to application strategies (e.g., flexible major assignment, targeting).

C.2 Personalized advising descriptions

C.2.1 Advising work-flow

We used a typical advising work-flow following the six-step structure described in the guidebook once we start to working with one student. Before that, after students added our advising account as friends, an administrative assistant confirmed her eligibility by verifying her Exam ID and School ID (in 2016, we could only verify school ID). The the assistant created a chat group for each student consisting with three people: the treated student, one advisor, and the assistant. In 2017, students had to complete a short survey to gain the eligibility (In 2016, we asked about individual information, such as track, CEE scores, preferences, through conversations).

- **Step 1.** A student (e.g., Alice) inputs her background information, including track, CEE scores (and subject scores), eligibilities for special programs, preferences (e.g., location, college type, majors).
  - In 2016, we asked about the individual information through conversations.
  - In 2017, students should complete a short survey before the start of advising.
- **Step 2.** The advisor (e.g., Motalk) or the assistant sends the guidebook (PDF file) to Alice and asks her to read the guidebook.
  - In 2016, we confirmed that all the "treated" students received the printed guidebook from their schools.
- **Step 3.** Motalk provides score equating results to Alice.
  - In 2016, we asked students to compute their equated scores by themselves. We provided them with the crosswalk table of scores and rankings to reduce their search cost.
  - In 2017, this was automatically completed in the Stata program (in a Stata log file, Figure 2).

- **Step 4.** Motalk provides a short list of colleges to Alice (short list is used to reduce search costs and to focus a student's time on researching the targeted set of colleges).
  - In 2016, we asked students to complete the search for a short list of colleges by using the admissions data in the books (a few hundred pages) provided by Ningxia Department of Education. Colleges in these books are alphabetically that it imposes high search costs for students to compare between colleges.
  - In 2017, this was automatically completed (based on the administrative data we received and were granted permissions to use from Ningxia Department of Education, as well as students' preferences data).
- **Step 5.** Alice returns a much shortened list of colleges in each institutional tier of her interest.
  - In 2016, this was done through intensive conversations. Advisors walked through the initial short list and helped students add/delete colleges.
  - In 2017, students were encouraged to take some time to look at the official website (and other information) of each college they are interested in before making the decisions.
- **Step 6.** Motalk provides the predicted probabilities of each college.
  - In 2016, this was done using subjective evaluations or rules of thumb (e.g., using 0.05 s.d. or 0.1 s.d. as the threshold; depending on individual preferences).
  - In 2017, we provided the admissions probabilities that were predicted by our machine learning algorithm (random forest) for each college-major-list order for each students (Figure 3).
- **Step 7.** Motalk helps Alice to finalize her application plan.
  - In both 2016 and 2017, this process involved many conversations about choosing the final four choices, considering different strategies (e.g., targeting), special programs, and college-major trade-offs. The decision would be based on the predictions in Step 6.

- **Step 8.** Alice completes online application in the Department of Education's centralized system.
  - We kept sending nudge, reminders and tips until the end of the college application period.

C.2.2 *How does machine learning work?*

We apply machine learning and other (big) data assisted methods, together with new technology (e.g., online survey and data synchronization tools) to increase the one-on-one advising efficiency. These data-based methods reduced the advisor's (and/or the student's) work in several ways:

1. The input of background information (using online survey and data synchronization) [Step 1]
2. Automatic score equating (in *Stata*) [Step 3]
3. Constructing short list of colleges (in *Stata*) [Step 4]
4. Predictions of admissions probabilities (in *R Shiny* with built-in machine learning predictions) [Step 5]

The reduced time of the advisors could be used to increase the number of students they could provide service to, and that for students could be used to deeper understand the knowledge and strategies of college applications and to better collect and analyze information and data about colleges (and majors).

C.2.3 *Example*

Figure C.1 provides an example of the one on one advising in 2016, which was similar to the machine learning assisted advising in 2017 as shown in Figure A.3.

The conversations show some behavioral barriers that students had and how we helped them in the college choice process. In Panel A, the student asked whether Shandong University was beyond the range of "reach college" to apply to. The advisor asked the student to do the CEE score equating and asked for the scores in the past three years (564, 588, 588). After reviewing the admissions data, the advisor replied that it was appropriate to list Shandong University as her first choice. In Panel B, the student sent a message after about one month reporting that "Thank you for your advising. I have been admitted to Shandong University."

(a) Adjusting the reach college choice (b) Admitted to the reach college

Figure C.1: Example of the online individualized advising in 2016

Notes: This figure shows a typical case of our 1-on-1 advising. The conversations were at QQ, one of the two largest chat forms in China.

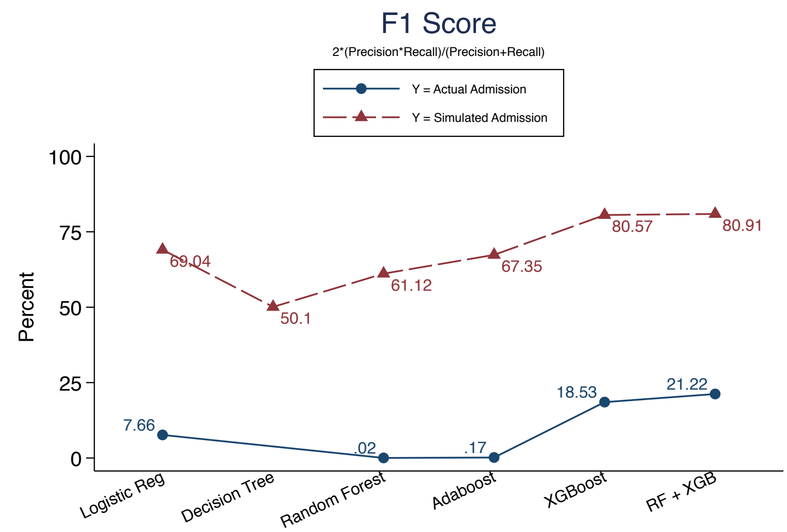
C.3 Prediction performance of different supervised learning models

The key element in the machine learning-assisted advising is to predict the college-major admissions probabilities for each student. As introduced in Subsection 3.2, I trained the prediction model using detailed applications and admissions data from the previous year and Random Forest. In a companion work in progress after the implementation of the experiment in 2017, I have tested other models that take much more time to develop. Random Forest overall shows good performance. I also tested XGBoost, an algorithm that has been dominating applied machine learning predictions using structured data (Chen and Guestrin, 2016).

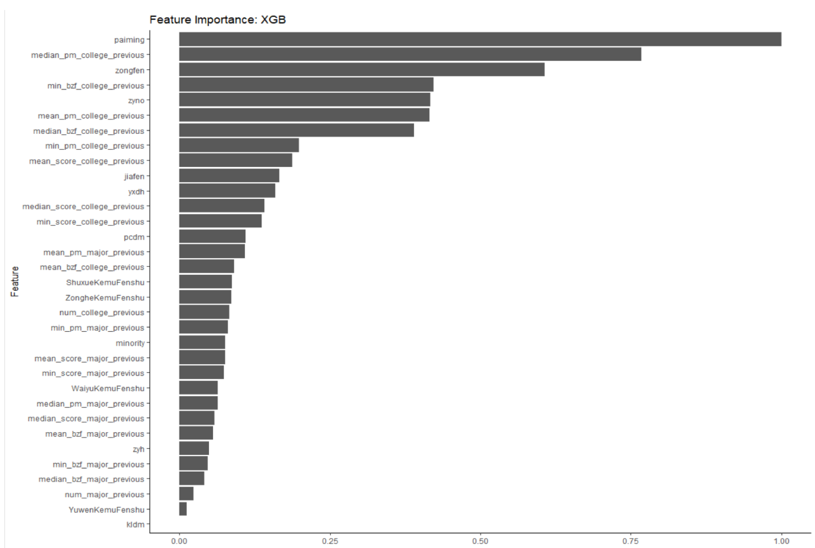
Panel A of Figure C.2 compares the prediction performance (based on F1 score) of six supervise learning models: Logistic regression, Decision Tree, Random Forest, Adaboost, XGBoost, and Random Forest + XGBoost. F1 score is evaluated using admissions outcomes in 2017. There are two important findings. First, boosting methods increase prediction accuracy than simpler methods including Random Forest and the combined application of Random Forest and XGBoost has the best performance among the six models evaluated. Unlike single models, boosting algorithms improve the prediction performance by training a sequence of models to minimize the errors from previous models and boosting high-performance models. XGBoost is an optimized version of gradient boosting that minimizes errors in sequential models.

Second, domain knowledge is required to build a reasonable prediction model. A data scientist might use the raw data where a student is admitted to a college-major and then train a model with very low F1 scores (blue line). However, in the Chinese college admissions system, students only receive one admission offer from the highest ranked colleges that their CEE scores qualify for. In the data, all the other lower-ranked colleges are coded "rejected." Therefore, it is necessary to create simulated admissions outcomes that assume students could receive multiple offers. This simulation is based on a good understanding of the admissions mechanisms (domain knowledge). The red line suggests that the models have much more accurate predictions for the simulated outcomes.

Panel B shows feature importance in the XGBoost predictions. Similar to the Random Forest prediction in Figure A.2, a student's ranking is the most important prediction of college-major admissions. XGBoost uses more information from college-major admissions results from the previous year than Random Forest.



(a) Model comparisons



(b) Feature importance in XGBoost

Figure C.2: Comparing different supervised learning models

C.4 Heterogeneity analysis using Causal Forest

For heterogeneity analysis, linear interaction model does not provide information on how covariates multiplicatively identify particular *individuals* who would benefit *most* from the individualized advising. To address this question, I apply cluster-robust Causal Forests to uncover the heterogeneous advising treatment effects (Athey and Imbens, 2016; Wager and Athey, 2018; Athey et al., 2019).[45] Similar to nearest neighbor methods, each single regression tree of the "forest" partitions the covariate space into small regions (leaves) and predicts the average outcome for individuals with particular covariates in that region. Causal Forests identify the conditional average treatment effect by aggregating a weighted average pair comparisons of neighbor observations (in small "leaves") from a set of trees built on random subsamples.

I grow the causal forest model using the covariates in the main effect estimation, including female, rural, STEM track, minority, repeater, and age. CEE score and randomization strata are also included to fit the outcome but not for building the forest.[46] For honesty, I randomly split half of the sample to build the model and the other half for estimation. Figure C.3 visualizes a single tree from the causal forest for estimating the effects of machine learning-assisted advising ($T1$) on college match index. According to the variable importance measures, STEM track (0.35), rural (0.18), and female (0.14) are the three most often chosen variables. Results are similar for other outcomes.

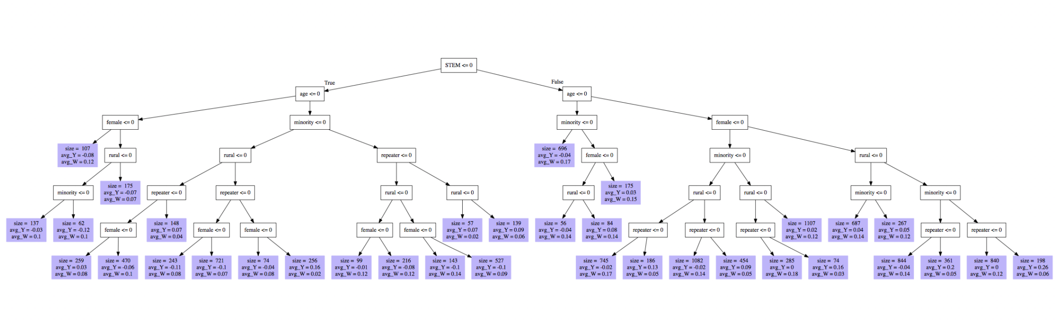
Figure C.4 shows the heterogeneous treatment effects of the machine learning-assisted advising ($T1$) at the individual level. Results on college admissions and college match are quite consistent. The conditional average treatment effects are similar to those estimated using the linear regressions in Table 3: 0.02 vs. 0.01 for college admissions and 0.035 vs. 0.022 for college match index. The majority of students would have a substantial positive effect. About one-fifth of the students in the sample are predicted to have a negative treatment effect; however, these negative effects

[45] Alternative machine learning approaches to estimating treatment effect heterogeneity include LASSO (Imai, Ratkovic et al., 2013), BART (Powers et al., 2018), and neural networks (Farrell, Liang and Misra, 2018). Causal forests method has known asymptotic properties and performs well in modeling high-dimensional nonlinear interactions between covariates.

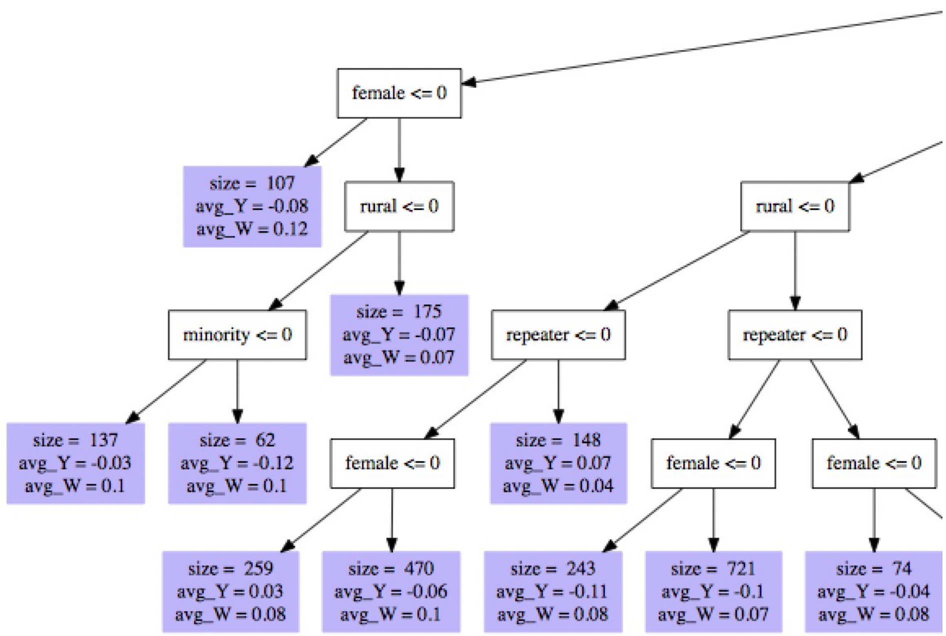
[46] Because CEE score has hundreds of potential splitting points that the tree method can be biased. Including randomization strata fixed effects ensures the internal validity of treatment effect estimation.

are not statistically different from zero. Students with estimates of individual treatment effects above the sample median have 4.9 percentage points (0.08 sd) larger treatment effects on college admissions (match) than students below the median. These statistically significant differences confirm the existence of treatment effect heterogeneity. Figure C.5 shows that the "business as usual" advising ($T2$) does not have detectable heterogeneous treatment effects with the conditional average treatment effects statistically indistinguishable from zero.

Table C.2 summarizes individual characteristics by the quartile of treatment effects estimated from the Causal Forests model. While there were smaller heterogeneous effects on college admissions, machine learning-assisted advising was predicted to have larger effects on college match for students from rural families. For example, 70.8% of students in the top quartile of treatment effects were rural students but only 36.8% of students in the bottom quartile were from rural families. Students with lower college entrance exam scores or who repeated the 12th grade were more responsive to the personalized advising. At the individual level, the Causal Forests model predicts that female, rural students who were in non-STEM track, non-repeater, non-minority, and at the normal age had the largest treatment effects. In contrast, female, urban students who were in non-STEM track, non-repeater, non-minority, and older than 19 years old had the smallest treatment effects. The existence of heterogeneous treatment effects of the individualized advising implies that providing access to advising is not enough to improve college-going outcomes for all. Complementary policies and interventions should be used to influence students to take the advising opportunities. To this goal, information about individual treatment effects will further help future work to target particular groups of students with different intervention designs.



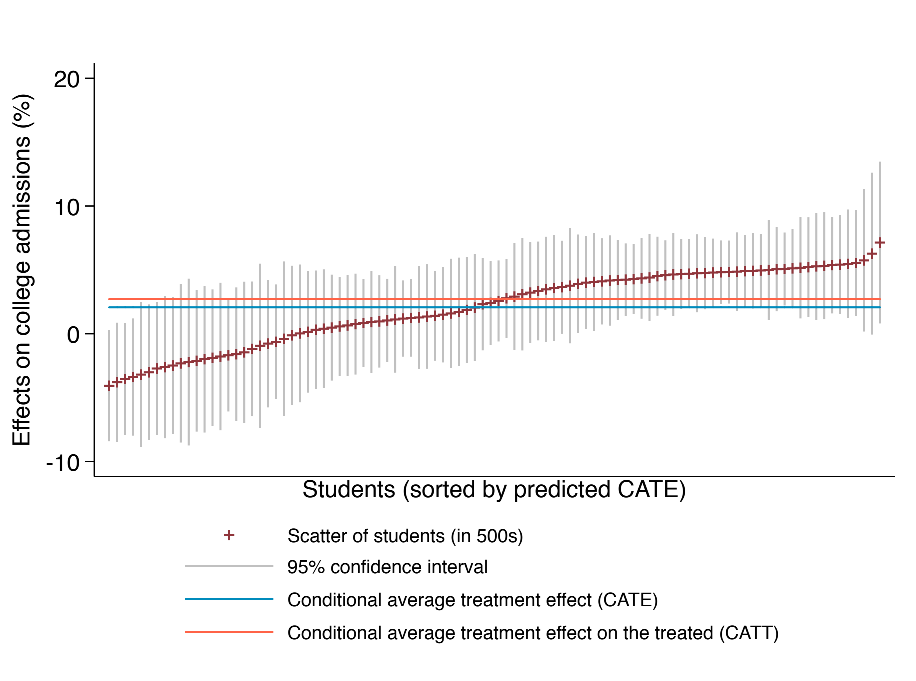
(a) Full



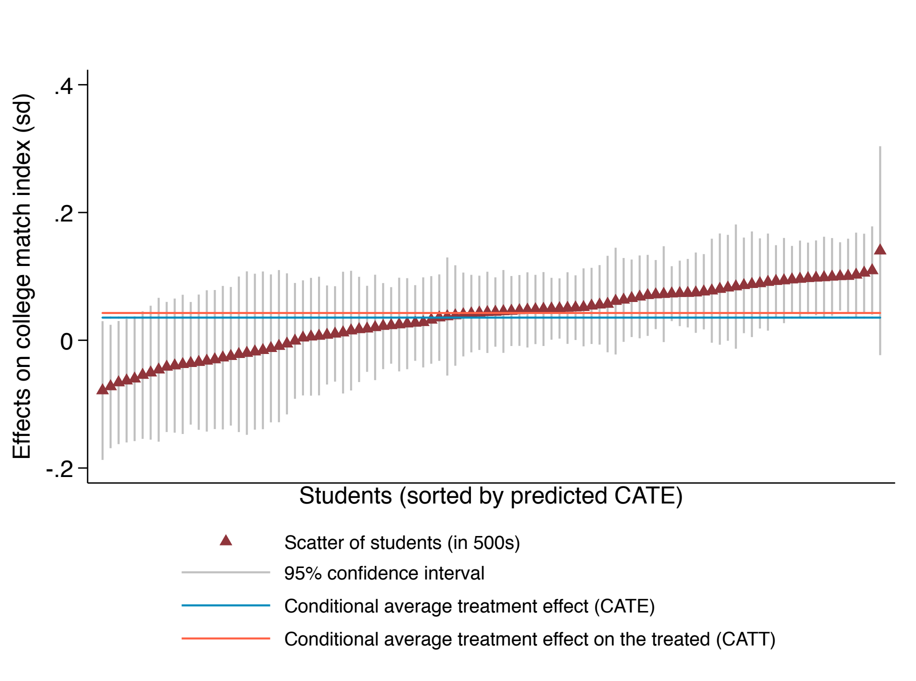
(b) Partial (leftmost part)

Figure C.3: A single tree for estimating the effects of machine learning-assisted advising on college match

Notes: This figure plots a single tree for estimating the effects of machine learning-assisted advising on college match using causal forests. Panel (a) shows the full tree and panel (b) shows the leftmost part. Each white box indicates a split and the corresponding covariate and each purple box indicates a leaf node. $size$ reports the sample size within each node. avg_Y and avg_W are the mean values for the outcome (college match index) and the treatment status within each node. The causal forests are estimate using 50% of the experimental sample (training data). **Age** is a dummy indicator for students younger than nineteen years old by the time of the college entrance exam.



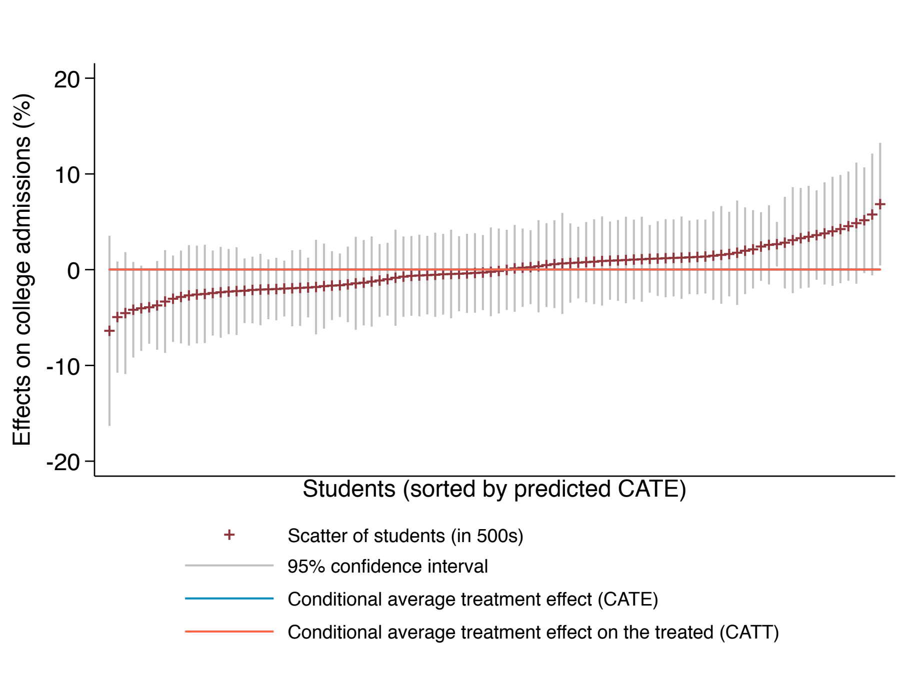
(a) Outcome = Admitted to any college



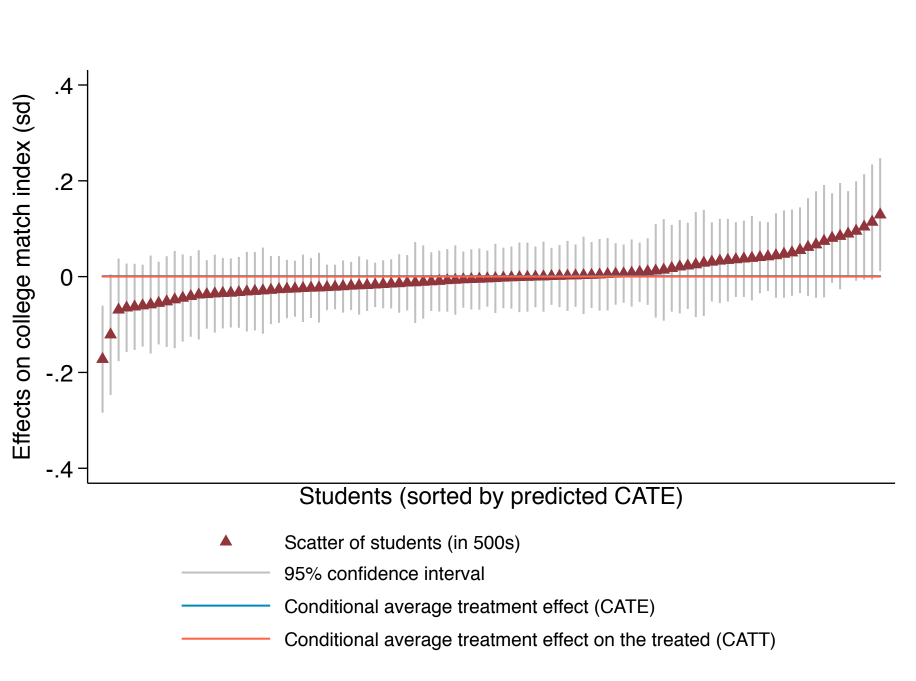
(b) Outcome = Admitted college quality

Figure C.4: Heterogeneous effects of machine learning-assisted advising by Causal Forests

Notes: This figure plots the heterogeneous effects of machine learning-assisted advising. Each scatter shows the average treatment effects of the 500 students within that scatter. In panel (a), the CATE for the full sample is 0.021 (s.e.=0.005) and the CATT is 0.027 (s.e.=0.005); the CATE for students with estimates of individual treatment effects above the sample median is 0.046 and that for students below the sample median is -0.004. In panel (b), the CATE for the full sample is 0.035 (s.e.=0.011) and the CATT is 0.043 (s.e.=0.010); the CATE for students with estimates of individual treatment effects above the sample median is 0.075 and that for students below the sample median is -0.006. The vertical lines are 95% confidence intervals.



(a) Outcome = Admitted to any college



(b) Outcome = Admitted college quality

Figure C.5: Heterogeneous effects of "business as usual" advising by Causal Forests

Notes: This figure plots the heterogeneous effects of "business as usual" advising. Each scatter shows the average treatment effects of the 500 students within that scatter. In panel (a), the CATE for the full sample is 0.001 (s.e.=0.009) and the CATT is 0.0001 (s.e.=0.009). In panel (b), the CATE for the full sample is 0.0002 (s.e.=0.005) and the CATT is 0.0002 (s.e.=0.005). The vertical lines are 95% confidence intervals.

Table C.2: Averages in covariates by the quartile of individual treatment effects from Causal Forests

	ML-assisted advising effects on admission				ML-assisted advising effects on college match index			
	Bottom quarter	Q2	Q3	Top quarter	Bottom quarter	Q2	Q3	Top quarter
Rural	0.547	0.513	0.538	0.530	**0.368**	**0.453**	**0.599**	**0.708**
	[0.498]	[0.500]	[0.499]	[0.499]	**[0.482]**	**[0.498]**	**[0.490]**	**[0.455]**
Female	0.575	0.551	0.536	0.520	0.530	0.543	0.547	0.561
	[0.494]	[0.497]	[0.499]	[0.500]	[0.499]	[0.498]	[0.498]	[0.496]
STEM	0.585	0.653	0.723	0.713	0.677	0.642	0.692	0.664
	[0.493]	[0.476]	[0.448]	[0.452]	[0.468]	[0.480]	[0.462]	[0.472]
Minority	0.312	0.321	0.338	0.344	0.313	0.283	0.273	0.447
	[0.463]	[0.467]	[0.473]	[0.475]	[0.464]	[0.450]	[0.445]	[0.497]
Repeater	0.195	0.158	0.222	0.185	0.081	0.185	0.228	0.266
	[0.396]	[0.365]	[0.416]	[0.388]	[0.272]	[0.389]	[0.420]	[0.442]
Age (<=19)	0.855	0.845	0.865	0.856	0.798	0.872	0.870	0.882
	[0.352]	[0.362]	[0.342]	[0.351]	[0.401]	[0.334]	[0.337]	[0.323]
CEE score (s.d.)	0.057	0.165	0.186	0.238	**0.551**	**0.138**	**-0.008**	**-0.035**
	[0.947]	[0.976]	[0.943]	[0.979]	**[0.995]**	**[0.950]**	**[0.869]**	**[0.921]**

Notes: This table shows the mean and standard deviation of each covariate by the quartile of individual treatment effects estimated from the Causal Forests model. The left panel is based on the effects of machine learning-assisted advising on college admissions and the right panel is based on the effects on college match index.