Improving College Choice in Centralized Admissions:

Experimental Evidence on the Importance of Precise Predictions*

Xiaoyang Ye

Brown University

*Xiaoyang Ye, xiaoyang.ye@brown.edu, Annenberg Institute for School Reform at Brown University. I gratefully acknowledge the essential guidance and support from my dissertation committee: Stephen DesJardins, Susan Dynarski, Brian Jacob, Brian McCall, and Jeffrey Smith. Thanks to editors Stephanie Cellini and Steven Hemelt as well as reviewers at EFP for their encouraging and insightful guidance on improving this paper significantly. I benefited from the helpful comments of Drew Anderson, John Bound, Dom Brewer, Sarah Cannon, Ben Castleman, Shuangye Chen, Xiaoyu Chen, Amy Claessens, Connor Cole, Sean Corcoran, Will Doyle, Alex Eble, Avi Feller, Thomas Goldring, Joshua Goodman, Wei Ha, Daniel Hubbard, Rodney Hughes, Ozan Jaquette, Dan Kreisman, Michal Kurlaender, Helen Ladd, Jing Liu, Susanna Loeb, Mark Long, Prashant Loyalka, Xinwei Ma, Andrew McEachin, Isaac McFarlin, Kata Mihaly, Karthik Muralidharan, Richard Murnane, Lindsay Page, Laura Perna, Natasha Pilkauskas, Xu Qin, Silvia Robles, Yang Song, Kevin Stange, Kaiyun Wang, Rong Wang, Shaoda Wang, Yaxin Wang, Noreen Webb, Christina Weiland, Po Yang, Lei Yu, Cheng Yuan, Liang Zhang, Jiusheng Zhu, and participants at Advances in Field Experiments, AEFP, AERA, APPAM, ASHE, NAEd Retreat, Peking University, RAND, Tsinghua University, University of Florida, University of Michigan, University of Washington, University of Wisconsin-Madison, and an anonymous reviewer at the ASHE conference. I particularly wish to thank Yanqing Ding (Peking University, the other co-PI), Wei Li (Ningxia University), Bingqing Dai and Wei Geng (Ningxia Department of Education) for their extraordinary effort in initiating the Bright Future of China Project, Xiaoyang Hu, Jing Liu, Le Kang, Fengming Lu, Liping Ma, Yujie Ma, Bingyu Sun, Yinduo Wu, and Jin Yang for their superb project leadership and management, as well as many research assistants and partners for the Bright Future of China Project. I also thank the China Institute for Educational Finance Research and the Institute of Economics of Education at Peking University, China Ministry of Education, and Ningxia Department of Education for providing the data used in this paper. This project would simply not have been possible without approval and collaboration from the Ningxia Department of Education. Yan Zhang has been a great help with data. 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, and the University of Michigan (Rackham, School of Education, International Institute, Center for Chinese Studies). The research was approved by the University of Michigan IRB (HUM00127559). All errors and omissions are my own.

Abstract

This paper provides the first experimental evidence of how admission outcomes in centralized systems depend on strategic college choice behaviors. Centralized college admissions simplify the application process and reduce students' informational barriers. However, such systems also reward informed and strategic college choices. In particular, centralized admissions can be difficult to navigate because they require students to understand how application portfolios and placement priorities map to admission probabilities. Using administrative data from one of the poorest provinces in China, I document that students made undermatched college choices that correlated with inaccurate predictions of admission probabilities. I then implemented a large-scale randomized experiment (N=32,834) to provide treated students with either (a) an application guidebook or (b) a guidebook plus a school workshop. Results suggest that informing students on choosing colleges and majors based on precise predictions of admission probabilities can effectively improve student-college academic match by 0.1 to 0.2 standard deviations among compliers without substantially changing their college-major preferences.

Keywords: College choice, behavioral intervention, centralized college admissions, field experiment

Running Head: Improving College Choice in Centralized Admissions

1. Introduction

Inequality in college access and match persists. Globally, low-income and minority students, facing various barriers at every stage of their educational pipeline, are much less likely to attend college and particularly selective institutions (Holsinger and Jacob, 2009; Bailey and Dynarski, 2011; Li et al., 2015). In recent years, the complex transition from high school to college has been increasingly recognized as an important barrier for students, especially those from disadvantaged backgrounds (Lavecchia et al., 2016; Page and Scott-Clayton, 2016; Dynarski et al., 2022). Even when low-income students reach the college choice stage, they are more likely than their high-income peers to apply to and enroll in colleges that are not matched to their academic achievements. That is, they undermatch (Bowen et al., 2009; Hoxby and Avery, 2013; Smith et al., 2013; Dillon and Smith, 2017). Undermatched college choice significantly lowers a student's chances of college and career success (Howell and Pender, 2016; Dillon and Smith, 2020; Kang and Torres, 2018; Ovink et al., 2018; Li et al., 2022). The undermatch problem for disadvantaged students is prevalent not only in the U.S. but also in Chile (Hastings et al., 2018), England (Campbell et al., 2022), Russia (Prakhov and Sergienko, 2017), China (Loyalka et al., 2017), and many other countries (Altmejd et al., 2022).

During the past decade, behavioral interventions, including provision of light-touch information and intensive personalized advising, have been widely proposed and implemented as promising policy tools to help students navigate the complex transition from high school to college.[1] The rapidly growing literature pertaining to college choice interventions focuses on decentralized admission systems such as in the U.S. and Canada, however, little is known about what works in other contexts.[2] Many countries use centralized college admission systems with mandatory entrance exams and a simplified application process, in which information and simplification in the

[1] Recent summaries include Thaler and Sunstein (2008), White House (2014), Castleman et al. (2015b), Lavecchia et al. (2016), Page and Scott-Clayton (2016), Castleman (2017), French and Oreopoulos (2017), Damgaard and Nielsen (2018), and Dynarski et al., 2022.

[2] A small body of literature has examined light-touch information about the benefits of college and major, financial aid, role model, and the importance of test scores. (Dinkelman and Martínez A, 2014; Hastings et al., 2018; Herber, 2018; Peter et al., 2018; Bonilla-Mejía et al., 2019).

application process is no longer the primary barrier. Instead, centralized admissions may reward informed, strategic applications based on precise predictions of admission probabilities and penalize application mistakes (Kapor et al., 2020; Larroucau et al., 2021).[3] Whether and how strategic college choices affects admission outcomes in centralized systems remains as an open question.

In this paper, I present one of the first experimental studies on college choice behaviors and admission outcomes of low-income students in a centralized college admissions system. I focus on students' strategic choices about which colleges and majors to apply to. Specifically, using administrative data from one of the poorest Chinese provinces (Ningxia), I document that - as is true in decentralized admissions - academic undermatch is also prevalent in centralized admissions.I examined a set of college-choice strategies and preferences constructed from students' detailed college application data. Among these measures, the most important predictors of college match were the *targeting strategies* for choosing an appropriate set of colleges and majors based on precise predictions of admission probabilities.[4] These targeting strategies resemble the expert advice that high-income American students often use in their college choices, as described by Hoxby and Turner (2013).

I then ask whether informing students about precise predictions of college admissions, without changing their college-major preferences, could improve these students' college application sets and admission outcomes in the Chinese centralized system. Because students have heterogeneous preferences for different types of colleges and majors,[5] a college choice intervention is only desirable if it helps students apply to the programs they prefer and get admitted by better academically match ones. Collaborating with the local government and high schools in Ningxia province, I conducted a

[3]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 (Neilson, 2019). In line with the growing adoption of centralized admissions in K-12 school choice, many U.S. 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.

[4]Kapor et al. (2020) report quite similar problems in the centralized school choice setting in New Haven, CT. Mulhern (2021) shows that, in the U.S. decentralized admissions, students' college choices are substantially affected by personalized admission information.

[5]For example, the "College Search" section of the College Board has ten filters: test scores & selectivity, type of school (2-year or 4-year, public or private, size, sing-sex or coed, religious affiliation), location, campus & housing, majors & learning environment, sports & activities, academic credit, paying, additional support programs, and diversity.

large-scale randomized controlled trial (RCT) with high school graduates of the 2016 graduation cohort. In a college choice advising program called the *Bright Future of China Project*, I designed a comprehensive college application guidebook.[6] The focus of the guidebook differed from existing literature by focusing on teaching students to make rational predictions of college admission probabilities and reduce application mistakes, regardless of their college-major preferences.

In a stratified school-level randomized experiment, I randomly assigned 32,834 high school graduates in Ningxia to either a control group or one of the two treatment groups that received either (a) the guidebook or (b) the guidebook and a school workshop during the 5-day college application period (see summary in Table 1). I found that these precise prediction-based interventions - provided in a short time with low costs - substantially improved college application behaviors and admission outcomes. A combination of guidebook and workshop produced larger treatment-on-the-treated effects (about 0.15 to 0.23 standard deviations in the college match index).

I answered two additional questions. First, the experimental evidence shows that we can help improve college admission outcomes by guiding students to use prediction-based *targeting strategies* without substantially changing their other preferences such as tuition, admission quotas, special programs, and majors. Second, the potential general equilibrium effects would be minimal and disadvantaged students are still likely to benefit from interventions even when every student can accurately estimate admission probabilities. This finding is consistent with recent work by Wang et al. (2022), which suggests that a de-biasing intervention would substantially improve the welfare of the disadvantaged students without hurting the advantaged ones.

This paper contributes to a growing literature on the effectiveness of behavioral interventions for the complex transition from high school to college (see the recent summaries in White House, 2014; Page and Scott-Clayton, 2016; J-PAL, 2018; Dynarski et al., 2022). The existing literature has

[6] Our research team named this project before knowing that the College Board has a similar program with a similar name ("Big Future"). Apparently, we all hope to help students gain bright/big futures. A few light-touch interventions might be less effective in centralized systems than in decentralized systems, such as reminders (because students receive a series of text messages from the Department of Education), application fee waiver (because students already pay for the very low college entrance exam testing and application fee), information/nudge/assistance of the application procedure (because centralized systems are simple and straightforward), information about college return (because almost all students are motivated to attend college), and information about college cost (because the information is centrally provided by the Department of Education).

primarily focused on students in developed countries and in decentralized systems (e.g., the United States and Canada). Very limited evidence has been available about what works to improve college decisions in centralized systems or in developing countries.[7] Centralized admission systems are widespread across countries in both K-12 and higher education. While centralization streamlines and simplifies the application process, it may require strategies and sophistication in decision-making, so that one would expect to see differences in the effectiveness of existing behavioral interventions (Pathak and Sönmez, 2013; Chen and Kesten, 2017). This paper provides new evidence about student-college academic undermatch and its potential sources regarding strategic application behaviors in centralized admissions.

This paper also presents experimental evidence on the impacts of providing application assistance with making precise predictions of college admissions as opposed to information provision and application simplification in the context of the largest centralized college admissions market in the world. The intervention designs in this paper build on many prominent approaches in decentralized systems, including information provision (Hoxby and Turner, 2013; Goodman, 2016; Peter and Zambre, 2017; Herber, 2018; Evans and Boatman, 2019), advising/counseling (Bettinger et al., 2012; Castleman et al., 2015a; Carruthers and Fox, 2016; Carrell and Sacerdote, 2017; Oreopoulos et al., 2017; Page et al., 2017; Castleman and Goodman, 2018; Gurantz et al., 2019), and school workshops or services (Oreopoulos and Ford, 2016; Bowman et al., 2018; Bettinger and Evans, 2019). This paper demonstrates an effective researcher-initiated, problem-solving intervention approach that focuses on a core component of college choice advising - informing students on making precise predictions of college admission probabilities.

The finding of the effectiveness of strategized college application interventions such as guidebooks and workshops is also consistent with recent literature on providing personalized information on admission probabilities in both decentralized and centralized systems. Arteaga et al. (2022) show evidence that beliefs about admission chances shape choice outcomes by influencing

[7]Experimental evidence on centralized admissions is limited, with few exceptions such as Hastings et al. (2018), Peter et al. (2018), Arteaga et al. (2022) in higher education, and Corcoran et al. (2018) and Arteaga et al. (2022) in the U.S. school choice. Many U.S. college admission officers and high school counselors advocate for a centralized system to simplify college choice, see a 2014 Washington Post article “What if Google ran the college application process?”

the ways that applicants search for schools. Their experiments in Chile and the United States demonstrated that providing live feedback on admission chances helps applicants search more effectively. Mulhern (2021) showed that personalized admission information that U.S. high school students received from the Naviance shifted applications and attendance to colleges for which students could observe information on schoolmates' admission experiences. These prediction-based interventions largely reduce students' college choice barriers such as uncertainty and present bias (Dynarski et al., 2021). This paper suggests that helping students to better analyze information and to make individualized, precise predictions is an important space for relatively low-cost policy interventions that can substantially improve their educational choices and outcomes at scale.

2. Background

2.1. Centralized College Applications and Admissions in China

This paper studies college choice in centralized college application and admission systems. China has the world's largest centralized college admission system that operates at the provincial level. The process begins with the administration of the annual national College Entrance Examination (hereafter CEE) in early June. The CEE scores are the sole criteria to rank students' priorities in college admissions. High school seniors take four subject exams: Mathematics, Chinese, English, and track-specific composite. Students choose either the STEM track, with composite exam items in physics, chemistry, and biology, or the non-STEM track, with composite exam items in history, social studies, and geography.

Next, all Chinese colleges allocate their college-major admission quotas to each province. All the information is centrally provided to students by the provincial Department of Education in mid-June. Shortly after, in late June, students learn their CEE scores and then submit their college application lists to the provincial Department of Education. Each student lists four to ten colleges for each institutional tier, varying across provinces-tiers; they rank colleges as well as majors within each college (typically six majors for each college). The submission process is simple in that students only need to type in the college and major codes into the online system.[8]

College admissions then proceed by institutional tier. Tier 1 includes the nation's elite colleges. Tier 2 and Tier 3 consist of non-elite public and private 4-year colleges, respectively. Tier 4 includes 3-year vocational colleges, which resemble community colleges in the United States. Tier 1 and Tier 2 colleges are selective and admit the top 30%-40% of applicants. Tier 3 and Tier 4 colleges are mostly open admissions and admit about 40% of the applicants who are relatively lower achieving. A student's application eligibility is limited to colleges in certain tiers based on their CEE score and the tier-specific admission cutoff scores, determined by the total number of spots and the distribution of the CEE scores within the province. Students with CEE scores above the Tier 1 cutoff are allowed to apply to colleges in all tiers. Students with CEE scores between the Tier 1 and Tier 2 cutoffs are only allowed to apply to Tiers 2-4 colleges, but not Tier 1 colleges, and so on.[9]

Based on their CEE scores, each student is matched with at most one college-major in their college application list through a predetermined matching mechanism - the Deferred Acceptance mechanism (Chen and Kesten, 2017). Each student receives a single take-it-or-leave-it admission offer for the college-major to which they are matched. Each year, about 20% of the CEE takers are not admitted by a college, including 10% of the CEE takers do not apply to any colleges and the other 10% apply but are rejected. If students decline their offers or do not receive an offer, they must wait until the following year to retake the CEE and participate in the college application and admission process again. The alternative is to enter the job market or to enroll in a foreign college.[10]

2.2. The Importance of Precise Predictions in Centralized College Admissions

College choice is a complicated decision for students, whether they apply through a decentralized or centralized system. Conventional college choice models (e.g., Manski and Wise, 1983; Long, 2004; Perna, 2006; Jacob et al., 2018) assume that a rational and forward-looking college

[8]In Appendix Subsection C.1, I show and explain in detail a typical college application form in China.

[9]Within each tier, there might be additional special admission programs that allow the eligible students to submit a separate application list. Special admissions include race- and income-based affirmative action programs, and early admissions for selected majors.

[10]Very few students decide to attend college outside China after taking the CEE. Those who aim to study abroad usually do not take the CEE and most of them have already made the enrollment decisions before the CEE in June.

applicant chooses from a feasible set of colleges and selects one that maximizes their expected utility.[11] The benefits of college include both "monetary" human capital returns and "non-monetary" preferences such as selectivity, college type, cost, distance, and consumption amenities (Jacob et al., 2018; Ovink et al., 2018). Therefore, students need to search for a large amount of college and major information and based on that information, apply to a list of colleges and majors that best fit their preferences. Recent literature has documented that students are boundedly rational: limited access to information drives undermatch outcomes, and providing information and guidance reduces inequality in college admissions (see summaries in Lavecchia et al., 2016; Page and Scott-Clayton, 2016; Castleman, 2017; Dynarski et al., 2022).

Moreover, college applicants are often highly sophisticated information processors with heterogeneous beliefs about admission probability (Bénabou and Tirole, 2016; Kapor et al., 2020).[12] Incorrect predictions of the admission probability will lead to mistakes in college applications. One extreme undermatch example is that, if a student applies to colleges in all of which she has very low admission chances, she might be rejected by all of them. Hoxby and Avery (2013) note that the expert advice concerning college applications in the U.S. decentralized system is to apply to several "peer" colleges, a few "reach" colleges, and a couple of "safety" colleges, where the types of colleges are defined by predicted admission probabilities.

In this paper, I focus on the importance of precise predictions in centralized college admission systems. Compared with decentralized admissions, centralized admissions are simplified and less costly.[13] In addition to taking the required, often nationally centralized college entrance exam, a

[11]College choice, in general, includes several stages. For example, Hossler and Gallagher (1987) propose a three-phase model (predisposition, search, and choice); DesJardins et al. (2006) jointly model the application, admission, financial aid determination, and enrollment decision process. This paper focuses on a student's decisions about which colleges to apply to.

[12]Avery and Hoxby (2004) find that students have different behavioral responses to what might objectively be viewed as similar dollar amount changes in the costs and benefits of college attendance. It can be viewed as a framing effect or nudge (Thaler and Sunstein, 2008; Benkert and Netzer, 2018), but it can also result from lack of knowledge to fully understand the real meaning of various forms of financial aid when they are labeled "grant" or "scholarship," and whether they are front-loaded.

[13]The expensive testing and applications in decentralized admissions may prevent students from applying to many colleges. Chade et al. (2014) note that a median American high school student applies to three colleges. Pallais (2015) finds that students strongly respond to an extra free college application (6$) in the SAT. Regarding the increased number of free ACT score reports available to low-income students, Hurwitz et al. (2017) also find positive effects on college attendance and degree completion of free SAT reports.

student only needs to submit their rank-ordered list of colleges and majors. They can apply to many colleges simultaneously and submit their applications online at minimal or zero cost. Centralized systems can also provide all the information relevant to a student's college applications, which addresses the common informational barriers in decentralized systems.

Although the application process is simplified and information is accessible, students in centralized admissions still face behavioral barriers that emphasize the role of strategic behaviors. The "expert advice" described by Hoxby and Avery (2013) also works in centralized systems: Assuming a student wants to be admitted to the highest quality or highest ranking college within the set of options that fit their preferences, applying to a mix of reach, match, and safety colleges is a reasonable strategy to maximize their expected admission outcomes. Conditional on a student's non-academic preferences, the expected admission outcome is a weighted average of each college's expected admission return (= college quality $\times$ admission probability) in a student's application portfolio.[14] This mixed-type strategy, which is widely used by students from high-income families, tends to maximize the expected admission outcomes of an application portfolio while maintaining a low risk of being rejected by all applied colleges.[15]

This mixed-type strategy is reasonable when further considering several common institutional features of centralized college admissions. First, students can not apply to all colleges and majors. In nearly all of the centralized college admission systems, students can only apply to a limited number of colleges (Neilson, 2019).[16] Assuming students correctly identify the types of colleges based on predicted admission probabilities, applying to a mix of different types of colleges could potentially improve their expected admission outcomes. For example, compared with applying to only one type of colleges, e.g., match colleges, which will most likely result in a student-college academic

[14]Because admission systems differ in whether they require students to apply to majors simultaneously, I focus on "applying to colleges" but the discussions can be easily applied to college-and-major choice.

[15]However, students with diverse risk preferences may adjust their use of this strategy. For example, students who are risk-loving and are willing to bear the consequences of not being admitted by any college would choose to apply to more reach colleges, while students who are risk-averse or have strong preferences for specific majors may only consider match or safety colleges.

[16]If the application list length is unrestricted, students can apply to all colleges and rank them based on their preference orders. However, when the list length is restricted, students have to strategically select and rank a subset of colleges.

match outcome, adding a safety college will largely minimize the chances of being rejected by all of the applied colleges; adding a reach college will increase the chances of entering a higher-quality college than match colleges.[17]

However, as students may have diverse preferences, the worth of applying to reach colleges depends on the expected returns of being admitted to a reach college and the cost of moving one application slot away from a match college. If students are risk-loving and highly value the increased quality of reach colleges than match or safety colleges and can take the consequences of being rejected by all colleges, applying only to reach colleges would be a preferred strategy for those students. Even in this case, having a safety or match college in the application portfolio would help largely reduce the rejection risk of the overall admission portfolios. Similarly, for students who are risk-averse and thus only apply to safety colleges, adding a match or reach college would increase the overall expected return without largely affecting the rejection risk of the admission portfolios.

Second, in many centralized systems, students will be matched to only one college based on the rank order of those colleges in their application list. Even though students can apply to as many colleges as possible that the selection of a subset of colleges is not needed, ranking colleges is still a critical decision for students. Conditional on that students are indifferent among a set of colleges that fit their preferences, the rank order should be primarily based on predicted admission probabilities. If a student ranks safety colleges before match colleges, they are very unlikely to be admitted by match colleges ranked lower in the list and thus likely to have an undermatched admission outcome. They also waste those application spots that can be used for applying to another match college to improve the expected admission outcomes of their whole application portfolios.[18] Misunderstanding the difference between unconditional and conditional probabilities often results

[17] If a student only applies to safety colleges, the admission probability is high but the college quality is low. If a student only applies to reach colleges, the college quality is high but the admission probability is low. The expected return as a product of probability and quality would be lower than a mixed application portfolio.

[18] This argument can be expressed in a sequentially conditional probability problem. Suppose college A has an unconditional admission probability of 0.5 and a quality index of 100, the expected return will be $0.5 \times 100 = 50$ if we rank it in the first choice. The expected return will be lower if we rank it in the second choice, which depends on the probability of the new first choice. If the new first choice is a reach college, e.g., with a probability of 0.05, the conditional expected return of college A only slightly changes to $(1 - 0.05) \times 0.05 \times 100 = 47.5$; however, if the new first choice is a safety college, e.g., with a probability of 0.95, the conditional expected return of college A dramatically reduces to $(1 - 0.95) \times 0.05 \times 100 = 2.5$.

in inappropriate ordering of the applied colleges.

Therefore, completing a reasonable college application requires a student to select a mixed set of colleges and rank them in an appropriate order as described above. The basis for doing so is to (a) precisely predict *ex ante* admission probabilities unconditionally for each college-major option and then (b) predict the admission outcomes conditional on different application portfolios. Students need to understand the college admission mechanism, access reliable sources of data on college admissions, and conduct educated predictions. They can make costly application mistakes if they incorrectly predict admission chances due to a lack of sophistication, a misunderstanding of the admission policies, or a limited ability to conduct "big data" analytics (Pathak and Sönmez, 2008; Kapor et al., 2020). The next section provides descriptive evidence on the importance of precise predictions of college admission probabilities in centralized systems.

3. Data, Variables, and Descriptive Evidence

3.1. Research Site and Data Sources

I empirically examine college application behaviors and admission outcomes using data from Ningxia province, officially the Ningxia Hui Autonomous Region, one of the poorest provinces in China.[19] As Chinese college admissions are centralized at the province level, such that students only compete for college-major spots with peer applicants within the same province, Ningxia offers a unique opportunity to focus on low-income students in a typical centralized college admission system. Each year, nearly all high school graduates (about 60,000) take the CEE.[20] About 90% of the exam takers apply to college and 85% are admitted to college. Fewer than 10% are admitted to elite colleges.

I used large-scale student-level administrative data for the universe of high school graduates

[19] Appendix Figure A.1 shows the geographic location of Ningxia. 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 Ningxia's 6 million population are under the extreme poverty line due to earning less than $1 a day.

[20] This is a highly selected population of "lucky" students who have overcome all the barriers from birth to grade 12. Nationally, only about 40% of a birth cohort (18 million) could reach the stage of college application.

in Ningxia in 2016. The data were provided by the Ningxia Department of Education and the Ningxia Education Examination Board, the provincial centralized administration office of the CEE and college admissions. The confidential student-level data are from four separate sources: (a) CEE registration data that include student demographic information, high school attendance records, and low-stakes graduation test scores; (b) CEE score data; (c) college application data that include all of the rank-order application lists that students submitted to the Ningxia Education Examination Board; and (d) college admission data that include the admission results for all students who have submitted their applications. The analytical data were de-identified.

3.2. Measuring College Admission Outcomes

During the past decade, the student-college academic match has drawn concern from education researchers and policymakers in many countries. In the United States where the undermatch literature emerged, it is widely believed that approximately 20%-70% of American high school graduates undermatch, though estimates vary across data, samples, and methods. Researchers have used various definitions of academic match based on data availability or specific research questions (see summary discussions in Rodriguez, 2015; House, 2017). Existing literature includes comparisons of student academic credentials with college selectivity (Roderick et al., 2008; Bowen et al., 2009; Smith et al., 2013), student ability percentile with enrollment weighted college quality percentile (Dillon and Smith, 2017), and comparing students' SAT/ACT scores with colleges' incoming freshman cohort median scores (Hoxby and Avery, 2013).

Extending the existing literature, I measure college access and match in centralized college admissions on both extensive (access) and intensive (match) margins. The *extensive* margin measure is a dichotomous indicator of college admissions that equals 1 if a student is admitted to a college, otherwise 0. For the *intensive* margins, I consider several measures that jointly denote the college match results. I first constructed an "undermatch" indicator, which equals 1 if a student is admitted to a college with a peer median CEE score 0.25 standard deviations lower than their CEE score, or when the student is not admitted to any colleges.

Table 2 documents the full extent of academic undermatch in a typical centralized admission system in China. The rows represent students' CEE score quartiles, indicating the quality level to which a student had access. The columns show the college quality level to which a student was admitted to a college. The student-college match in centralized admissions is very similar to that in decentralized admissions, notably in the United States (see, for instance, Smith et al., 2013, Dillon and Smith, 2017). Students showed an assortative matching pattern such that 65.9% of students concentrated along the diagonal. However, about 25% of students were admitted to a college that was one quality level below the level to which they had access. The change in overmatch is not accordingly symmetric as noted by Dillon and Smith (2017). Overall, 9.1% of students ended up with overmatched colleges based on the quartile matrix. Using the undermatch indicator as discussed above, 28.63% of undermatched students who were admitted to a college with a median CEE score 0.25 standard deviation lower than their own CEE scores.[21]

In addition, I use college-level median, mean, and minimum CEE scores of the admitted students in the same year to measure contemporaneous college match quality.[22] Holding the CEE score constant, a negative difference in college peer quality - resulting from the same year's college admissions - means that a student has "wasted" their CEE score to be academically undermatched with that college. To minimize the potential bias of using the college admission results in a single year and a single province to denote college quality, and to compare results across years, I used two national college quality measures: standardized score and ranking percentile.[23] Using these five intensive margin measures of college match, I construct a single index using principal component

[21]The measure is slightly smaller than that (33.5%) from the five-percentile threshold as proposed in Hoxby and Avery (2013). The choice of 0.25 SD as a conservative threshold is based on the practical experience of college choice advising in China. Table B.1 shows that the results remain qualitatively unchanged using other thresholds. Throughout the paper, all the results using these various undermatch indicators remain consistent. The national average undermatch rate for those admitted by four-year colleges decreased from 30% in 2005 to 15% in 2011, mainly due to the change from the Boston mechanism to the Deferred Acceptance mechanism (Chen and Kesten, 2017).

[22]The results remain unchanged if I use leave-one-own-out scores. For students who are not admitted by any college, I assign the tier-specific lowest college median/mean score minus 0.2 SD as their "college median/mean score." Results are very consistent using different measures for this group of students.

[23]Using college admission data from 1996-2017 and administrative data on institutional resources for every college in China, I build a national college ranking of all Chinese colleges, which has been used to assist all Chines high school graduates in their college choices.

factor analysis as the primary college match measure.[24]

In all of the college access and match measures, low-income and minority students were substantially more likely to undermatch than advantaged students. This poverty gap persisted even when controlling for demographics, CEE scores, and high school fixed effects. As centralized college admissions depend on CEE scores and student applications, this result suggests that the gap in college match, conditional on having the same CEE scores, is solely driven by between-student differences in college application behaviors.

3.3. Measuring College Application Behaviors

With the unusual access to the data of the college application list of each student, I explored their college application behaviors. I categorized a series of strategy and preference measures based on the existing literature and my college advising field experiences, including six groups of measures: (a) targeting strategies, (b) general advice, (c) special programs, (d) tuition and quota, (e) geographic location, and (f) major choice. Each group includes several variables measuring a student's college and major applications. Appendix Subsection C.1 provides a detailed description of those measures, explaining the behavioral rationales of these college application behaviors. I briefly discuss the key measures below.

The focus of this paper is college choice based on precise predictions of college admission probabilities, which are summarized in *targeting strategies*. The key targeting strategies include equating CEE scores to percentiles, applying to colleges in the match tier using a targeted portfolio with a mix of different types of colleges, and ranking colleges in descending order. Most of the *targeting strategies* measures can be objectively evaluated and therefore we can describe whether students make such "mistakes." For example, as discussed in Section 2, while applying to different

[24]Because these measures increase with both matched and overmatched admission results, the label of "college match measure" indicates the comparison between match/overmatch and undermatch. Principal component analysis is for data reduction by creating an index that summarizes conceptually similar measures. The five college match measures are highly correlated with each other. Pairwise correlation coefficients range from 0.740 (college-level minimum CEE score and standardized score of college quality) to 0.998 (college-level mean CEE score and median CEE score). Factor loadings from the principal component factor analysis are 0.9875, 0.9878, 0.8718, 0.9513, and 0.9600, respectively. After rotation, the regression coefficients used to estimate individual factor index scores for each of these five measures are 0.218, 0.218, 0.192, 0.210, and 0.211. The scoring coefficients sum up to 1. Minor differences exist due to rounding.

types of colleges depend on individual preferences, not equating CEE scores or ranking colleges in wrong order is a costly mistake that students make. The remaining measures are generally "expert advice" or recommendations for students (Hoxby and Avery, 2013; Arteaga et al., 2022).[25]

Precise predictions require students to understand the underlying mechanisms of college admissions: Ranking percentile, not the raw score, matters. Many students (a 65% estimate in this paper) naively compared their CEE scores in the application year with college admission raw scores from the previous years, resulting in mistakes in identifying college types (i.e., reach, match, and safety colleges).[26] As the application data do not show whether students equated their CEE scores to percentiles, I inferred this behavior by calculating the gap between a student's CEE score and the equated admission score from the previous year of their most likely targeting college and then defined that a student had equated the CEE score in their application if the estimated gap is within 0.15 standard deviations.[27]

As described earlier, Chinese college admission process proceeds with tiers, and a student's eligibility to apply to specific tiers depends on their CEE score and the tier cutoff scores. A student with a high CEE score can apply to lower-tiered colleges, but those colleges are all academically undermatched for the student. Therefore, it is crucial that students correctly identify their match tiers as the highest possible selectivity tiers that they qualify for. I estimated that, in the analytical sample, more than 23% of students did not apply to colleges in their match tiers.

Finally, as introduced in Section 2, students can target their application portfolios to a mix of reach, match, and safety colleges to maximize their expected college admission outcomes. In

[25]The word "mistake" refers to those strategies that can be objectively evaluated and we believed that all students should follow the rules. For example, students make mistakes in ordering their applications even when they have heterogeneous preferences. Imagine students prefer a less selective college, they still have to rank their applied colleges in the reverse order of admission probabilities. It is a mistake to list a reach college after a safety college because the conditional admission probability of the reach college can be so close to zero that the student just wastes the application spot. The remaining measures are only suggestions to students and students can adopt them according to individual preferences.

[26]In 2018-2022, I provided online advising to more than 200,000 high school graduates across China (not in a randomized experimental sample). Students were most interested in score equating and targeting the match colleges.

[27]Students have a very low admission probability when the gap between their CEE score and the college's admission score is larger than 0.15 SD. Results are qualitatively unchanged when varying the radius cutoff or the definition of a targeting college. In the main analysis, targeting college is defined as the second-choice college in a student's match tier as students are likely to apply to a reach college in the first choice.

addition, they need to understand that the rank-order of colleges in their application lists matters. This means that they should rank reach colleges before match and safety colleges. It is very common for students to misunderstand conditional probability and make mistakes in the ordering of applied colleges, resulting from either incorrect predictions of the admission probability of each college or a misunderstanding of the assignment mechanism. Kapor et al. (2020) and Arteaga et al. (2022) find that this ranking mistake in the application list is also common in the U.S. centralized school choice and the Chilean centralized college choice.

Figure 1 presents descriptive analyses of students' college applications. It shows the distribution of students by the gap between their CEE scores and the median scores in colleges of their first choice and fourth (last) choice in the match tier. It clearly shows that some students acted as the *targeting strategies* describe: They applied to colleges with median admission scores close to their CEE scores and ranked their applications in an order such that the first choice aims higher than the last choice. However, though correctly centered, a large proportion of students applied to colleges to which they would be substantially undermatched (the right tail) or where they had a nearly zero chance of getting in (the left tail). In particular, students who applied to an undermatched college in their first choice (the left tail) would certainly be admitted. In contrast, students who applied to a reach college in their last choice (the right tail) as well as reach colleges in the first three choices would likely be rejected by all of their applications in this selectivity tier.

A final *targeting strategies* measure I considered is whether students applied to college-majors that did not have historical admission data because they were to recruit students in a province for the first time in an application year.[28] In this case, students need to infer/predict the admission data in previous years for these "new" college-majors using other information. They also have to take the risks of applying to these college-majors because the predictions can be inaccurate. However, if most students are risk-averse and do not apply to those colleges, it is a good opportunity for strategic students to gain a matched or overmatched admission. Table C.2 presents supporting evidence that

[28]The "new" college-majors are from two sources. First, colleges allocate their major-specific quotas in different provinces across the country and switch the majors in a given province. For example, only about 750 of the 2,900 colleges recruit students from Ningxia. Second, colleges consistently open new majors over years.

applying to the "new" college-majors is associated with an increased probability of college match. I included this strategy in the college application guidebook to nudge students to carefully consider this "invisible" opportunity, however, which depends on students' effort in making the inferences of admission probabilities when historical data are not available.

I constructed another two groups of measures of college application strategies. *General advice* strategies describe college application strategies that can be improved through light-touch application guidelines, such as the number of applied colleges, the number of applied majors, and flexible assignment. For example, in Chinese college admissions, 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 admission scores than that student's CEE score. If the student accepts flexible major assignment, then the college will assign her to a major that still has a spot (but that major may not be the one the student is interested in). Accepting "flexible assignment" does not affect how students choose and rank majors in their application lists.

Special programs strategies summarize whether a student applied to affirmative action programs, early admission programs, or teachers' education programs. Those programs vary greatly in quality and application eligibility, and some of which are considered to be opportunities for students with lower CEE scores to access higher-quality colleges. However, students may lack awareness of and information about those programs.

The next three groups of college choice behavior measures relate to individual preferences for colleges and majors. *Tuition and quota* preferences include measures of the median college tuition and mean quota of all the colleges that the student applied to. Low-income students may prefer low-tuition colleges, and risk-averse students may prefer colleges with larger admission quotas (Hoxby and Avery, 2013; Loyalka et al., 2017). *Geographic location* preferences include a variable of the percentage of colleges that a student applied to were out of their home province (excluding in economically advanced regions) as well as a variable of the percentage of colleges that a student applied to were in economically advanced regions. *Major choice* preferences summarize the share of majors that a student applies to in each major category.

3.4. The Potential for Behavioral Interventions

Table 3 shows the correlations between college application behaviors and admission outcomes, estimated from a linear regression. Each group of the strategy and preference measures is summarized as a standardized index using the principal component analysis. Appendix Table C.2 presents estimates for each item of application behaviors, and results are consistent with those principal component indices.

The regression results indicate that *targeting strategies* are the most important predictors of improved college access and match outcomes. In Column 4, for students with the same CEE scores, the same demographic characteristics, and attending the same high school, each 1 *SD* increase in the use of targeting strategies is associated with a 0.217 *SD* increase in college match quality, holding other strategies and preferences equal. Column 7 of Appendix Table C.2 shows that all the items in targeting strategies are statistically significantly correlated with college match. For example, conditional on making precise predictions as indicated in the other measures in targeting strategies, applying to a mix of reach, match, and safety colleges is associated with a 0.03 *SD* increase in college match; ranking the applied colleges in descending order based on predicted admission cutoffs is associated with a 0.08 *SD* increase in college match.

In addition, preferences for lower tuition or larger quotas are associated with reduced college match quality, while choosing out-of-province options increases the probability of attending a higher-quality college. However, the correlation coefficients are smaller than one-third of the coefficient on *targeting strategies*. Taken together, indices of general advice, special programs, and major choice are not statistically or economically significantly correlated with admission outcomes.

The descriptive results shown above demonstrate the importance of precise predictions of college admission probabilities. A large proportion of students may have informational and behavioral barriers that they do not use *targeting strategies* that would improve their admission outcomes. Disadvantaged students are more likely to have such problems and thus undermatch.

Helping students make better predictions seems to be the most important and promising element of a behavioral intervention for improving college access and match in centralized systems.

4. Experimental Design

4.1. Interventions: Informing Students on Precise Predictions of College Admissions

I used a large-scale randomized controlled trial to test whether informing students about making precise predictions of college admissions improves their college-choice decisions and college-going outcomes. Considering the actions an expert counselor or a very sophisticated student would take in their college-choice decisions, I prepared a comprehensive college-choice guidebook with a focus on precise predictions of college admissions. I then designed two school-based channels to deliver the same information: (a) an application guidebook, and (b) a guidebook plus a school workshop.

The intervention design in the *Bright Future of China Project* was initially built on the application strategies in Hoxby and Turner (2013)'s Expanding College Opportunities project.[29] It combines features of both informational interventions and individualized advising/nudging examined in a wide body of literature (see summaries in Page and Scott-Clayton, 2016; J-PAL, 2018). I focused exclusively on the instruction and learning of the sophisticated college-choice strategies during a very short time (5 days) when students applied to college.

I led an expert team to prepare the "How to Apply to College" guidebook. The team included professors and graduate students in the field of education policy, school counselors, and college admission officers in China. Using our expertise in advising college choice for more than a decade and conducting additional learning from many prominent sources,[30] our research team

[29]I do not incorporate the other interventions in Hoxby and Turner (2013) and other related studies including cost information, application fee waiver, and parent intervention. In the Chinese centralized admissions, students are provided with tuition information for every college-major, and institutional financial aid is rare. College application fees are low ($25 with exam fees included). Nearly all high school seniors take the college entrance exam. In low-income areas, average schooling level of parents is lower than junior high school, which makes using any written materials mailed to parents ineffective.

[30]We have learned greatly from some excellent resources in the U.S., such as MDRC's "In Search of a Match: A Guide for Helping Students Make Informed College Choices" and the College Board's Big Future program. Our

produced a comprehensive guidebook. The guidebook consisted of four main "course" modules: (a) searching for college information, (b) understanding admission policies, (c) equating CEE scores and predicting college admission probabilities, and (d) applying to an appropriate portfolio of colleges and majors. To supplement the main modules, I also used large-scale (and confidential) databases and reliable official information about colleges and majors.

In advising students about how to search for information, I provided a table that maps a list of recommended websites of college-major information (panel A of Figure C.2).[31] To assist students with major choice, I used the post-graduation employment data of the universe of Chinese college students from 2011 to 2014, a dataset with over 30 million observations, to show employment rate trends (panel B of Figure C.2). Lastly and most importantly, I provided detailed explanations of college admission policies and actionable strategies to generate an optimal portfolio of colleges based on precise predictions of college admission probabilities. Subsection C.2 provides detailed descriptions and sample pictures of the guidebook (Figure C.1).

The guidebook was not designed to change students' college and major preferences, as it was not tailored to individualized applications. The exception is that it did nudge students to apply to out-of-province colleges. Existing literature has documented that the "home bias" in college choice often limits high-quality college opportunities (Hoxby, 2000; Long, 2004; Hillman, 2016; Ovink et al., 2018). Our previous work suggests that the preference for in-province colleges results in large welfare losses (Kang et al., 2020) because Ningxia, as one of the poorest provinces, lacks high-quality colleges.

4.2. Treatment Arms: (a) Guidebook and (b) Guidebook plus School Workshop

With the assistance of the Ningxia Department of Education, I distributed the "How to Apply to College" guidebook to all the students in the "Guidebook" treatment schools through the school

---

research team carefully reviewed more than 200 Chinese websites and guidebooks that contained information about college entrance exams and college applications. I identified the most reliable and useful information that later was synthesized in the guidebook.

[31]There are various online sources available to Chinese students, but most of them are unreliable and contain mistakes. It is not easy for students to find the reliable sources of information about college applications and to understand how to navigate the sources to find the information they need.

administration. Students were informed that the guidebook was prepared by researchers at Peking University, the top college in China, and at Ningxia University, the top college in Ningxia. The guidebook was expected to help students gather information and facilitate their learning of the rules and principles necessary to make knowledgeable decisions for themselves.

To advance students' learning of the guidebook, I worked with local districts and high school leaders to plan and run accompanying school workshops in the "Guidebook plus School Workshop" treatment schools. To minimize the quality variations in workshop delivery, I selected a group of very knowledgeable experts - the guidebook editors - to give the workshops using the same slides and scripts at each school. Workshops were announced one month ahead of time and described as being provided by a joint research team from Peking University and Ningxia University. Each workshop lasted 3 hours and was moderated by the school principal or a vice principal. All the four "course" modules were covered in detail by the speakers and a question-and-answer session was included. Figure C.3 shows sample pictures of the workshops. In those "Guidebook plus School Workshop" schools, I distributed the booklet to students who attended the school workshop.

4.3. School-Level Randomization, Implementation, and Summary Statistics

The randomization was at the school level. As requested by the Ningxia Department of Education, I first randomly selected three cities out of the five prefecture cities in Ningxia. I implemented the experiment in all public high schools in these three cities, which resulted in 31 schools (out of the total 60 schools in Ningxia) in the experimental sample. I then created four strata for the three cities by dividing the capital city into two strata based on school quality.[32] Within each stratum, I randomly assigned three schools to receive the guidebook treatment, two schools to receive both the guidebook and the workshop, and the remaining three schools to not receive either treatment, serving as the control group.[33] Table 1 presents the experimental design.

The experimental sample included 32,834 students who graduated from 31 public high schools

[32]The reason is that the most selective high schools concentrate in the capital city. School quality is measured using confidential school finance data in 2013, the latest year of the data I obtained from the China Ministry of Education.

[33]The number slightly varies across strata due to rounding.

in 2016. Of these, 11,408 students were in 12 control schools and received no treatment. In mid-June, before students submitted their college applications, 12,823 students in 12 schools were provided with the guidebook (T1). The guidebooks were sent directly to each treatment school, and school administrators distributed them to individual students when they received their score reports.[34] Another 8,603 students in seven schools were provided with both the guidebook and the workshop (T2).[35] Workshops were held during 22-24 June, 2016, when students were starting to submit their college applications (the deadline was June 27). Both treatment arms were at a relatively low cost due to their scalability.[36]

I was unfortunately not able to identify an accurate, individual-level take-up of the school workshops because schools failed to track the "treated" students (attendees) due to the lack of incentive and organizational capacity in these high schools. According to the number of booklets distributed to students, the take-up in T1 schools was 98%, and approximately 42% of students (29%-56%, varying by school) in the T2 schools attended the workshop and received the guidebook. Students in T2 schools who did not attend the workshop did not receive the guidebook.

The summary statistics in Table B.2 indicate that the experimental sample is representative of the entire high school graduation cohort. Exceptions are that the experimental sample had a 0.11 *SD* higher average CEE score, a 6% lower minority student fraction, and a 7% higher rural student fraction. About 60% of students were from rural families, about 30% were minorities (mostly Muslims), and about 20% of college applicants repeated the 12th grade at least once. The average college admission rate was 84 percent.

Within the experimental sample, mean student characteristics differ slightly between groups because the randomization used school-level finance data in 2013. On average, T1 had more rural students, and relatedly, lower-achieving students. I classify high/low achieving students using the

[34]Students can check their CEE scores online, but they are required to receive a formal printed report.

[35]I initially randomized eight schools for the workshop. One workshop was not held due to the ineffective school organization. I coded that school in T1. Results do not change if I drop this school from the analysis.

[36]The average cost of the guidebook was approximately $5 per student, including a budgeted personnel cost of $20,000 in total, a printing cost of $2 for each booklet, and a delivery cost of $2 for each booklet. The average cost of providing a school workshop was $2,000 for personnel and traveling costs. Given the estimate from the field, each workshop had an average attendance of 500 students and thus the per-student cost was $4. The average cost could be further reduced by increasing the number of treated students.

selective (tier 1) college admission cutoff. However, controlling for strata fixed effects, these three groups are balanced in observed characteristics for schools (using both the 2013 finance data that were used for randomization and the 2016 sample student data; see Table B.3) and for students (for both the whole sample and the high-achieving sample in 2016; see Table B.4).

4.4. Impact Evaluation

I examined whether and how the interventions altered students' college-choice behaviors and how the changes in college-choice behaviors affected admission outcomes.[37] I used the following linear regression to estimate the intent-to-treat effects (ITT):

$$Y_{ij} = \beta_0 + \beta_1 * T1(guidebook)_j + \beta_2 * T2(guidebook + workshop)_j + X_i * \gamma + \delta_s + \varepsilon_{ij} \quad (1)$$

where Y_{ij} is the outcome for student i in school j of randomized stratum s. $T1_j$ and $T2_j$ are indicator variables for school j receiving the guidebook or the guidebook-workshop treatments, respectively. δ_s are strata fixed effects. X_i includes a set of student characteristics, including a student's CEE score and demographics (gender, race, age, STEM/non-STEM track, repeater) to account for group differences in college preferences. All standard errors are clustered at schools.

I addressed multiple hypothesis testing in several ways. I constructed the outcome measures closely following the literature. For example, the college admission outcomes were from different yet highly correlated perspectives, which jointly provided a complete picture of college access and match. I primarily aggregated the outcome measures to several single indices to minimize the potential multiple hypothesis testing bias. Additionally, I applied the method proposed by List et al. (2016) to confirm the robustness of results.

Another issue of the cluster randomized experiment is the relatively small number of clusters (schools), which may result in incorrect statistical hypothesis tests (e.g., in *p*-values) based on large number asymptotic properties. I used randomization inference to assess whether the observed treatment effects were likely to have been observed by chance even if treatment had no effect (Heß,

[37]The main college match measures were explored in Loyalka et al. (2017), which motivated the development of the *Bright Future of China Project*.

2017).[38] In the regression tables, I report p-values from 1,000 permutations.

According to anecdotal evidence, the take-up of T1 is more than twice that of T2. Thus, the treatment-on-the-treated (TOT) effects would be of policy interest as well. To approximately compute the TOT effects, we can use a Wald estimator to rescale the ITT effects by the take-up probabilities. Based on the homogeneous treatment effect assumption, the approximates provide a sense of the results if we could scale up the interventions through making the guidebook and workshop a mandatory part of the high school curriculum or counseling. However, we should be cautious as it might be reasonable to assume that the learning from the workshop spilled over from workshop participants to their classmates with whom they communicated. Future research could use a better research design to identify such TOT effects and spillover effects.

5. Results

5.1. Effects on College Application Behaviors

Table 4 shows that the interventions substantially altered students' college-choice behaviors with the main changes in the prediction-based targeting strategies. Column 1 uses the single principal-component factor index to summarize the ITT effects on college application behaviors.[39] The guidebook-workshop intervention statistically significantly and substantially improved college applications. On average, students in the guidebook-workshop treatment schools submitted college applications with a 0.167 *SD* ($p<0.05$) higher quality index, a nearly 100% increase from the control

[38]Randomization inference or permutation tests, as introduced by Ronald Fisher in 1935, can handle the inference problems with small samples where traditional large sample-based regressions produce incorrect results.Under randomization inference in experiments, the sample is fixed while the assignment to treatment or control groups can be seen as random. We obtain a test statistic (e.g., treatment effect estimate) from each realization of the assignment. We then compute the exact distribution of the test statistic from many different random assignments (i.e., permutation) under the sharp null hypothesis of no treatment effect. We use the conventional significance level (e.g., 0.05) to reject the null hypothesis.

[39]I used the same approach as used in constructing the college match index and each of the six application behavior indices. Factor loadings for the six application behavior indices from the principal component factor analysis are 0.472, 0.646, 0.385, -0.706, 0.828, and 0.367, respectively. After rotation, the regression coefficients used to estimate individual factor index scores (reported in column 1) for each of these six application behavior indices (reported in columns 2-7) are 0.224, 0.307, 0.183, -0.335, 0.393, and 0.174. The scoring coefficients sum up to 1. Minor differences exist due to rounding.

group mean. Given that the take-up is about 40%, the TOT effect is roughly 2.5 times as large.[40] The guidebook alone may have improved applications, but the estimate is imprecise and of a smaller magnitude.

In columns 2-7, I test the strategy and preference groups separately. Consistent with the descriptive results in Table 3 - showing that the targeting strategies are the most important factors driving college match - both the guidebook and the guidebook-workshop interventions statistically significantly and substantially improved students' use of *targeting strategies* with a more than 100% increase from control group mean (column 3). In contrast, the interventions did not meaningfully affect students' other strategic application behaviors, neither *general advice* nor *special programs* strategies. Except for geographic locations, students improved their college applications (as measured by the quality index) not at a cost of changing their preferences for colleges and majors.

Table 5 reports estimates for each college choice behavior item. These estimates confirm that the effects well aligned with the focus of precise predictions in the intervention designs. Changes in *targeting strategies* were not from just one or two items by chance but from improvements in all elements that contributed to an optimal college application. The ITT effects show that students were more than 10% (3 to 4 percentage points) more likely to apply to a mix of reach, match, and safety colleges (column 7) and to rank these colleges in descending order of predicted admission cutoffs (column 6). Though imprecisely estimated, students also appeared to be more likely to equate their CEE scores (indicated by the estimated gap in column 4) and more likely to apply to colleges in the institutional tiers that matched their CEE scores (column 5). With improved prediction skills, they were also more than 20% more likely (statistically insignificantly) to apply to colleges without admission data in the prior year (column 8). Taken together, these results consistently indicate that treated students have substantially improved their precise predictions.

In addition to helping students better predict college admission probabilities, the interventions

[40] Potential general equilibrium effects and heterogeneous treatment effects might lead that the TOT effect is different from the ATE of the whole sample. As I will discuss later, while I found heterogeneities in the treatment effects, the general equilibrium effect would be minimal.

also largely shifted students from colleges in Ningxia or neighboring low-income regions to out-of-province colleges, especially in economically developed regions. While the guidebook and workshop presented the same information regarding geographic preference, the differences in magnitude and statistical significance between the two treatment groups might be from the increased nudge and salience brought to students during the school workshop.

I should note that the college application behaviors characterized in this paper did not fully capture how students made their choices, given that they could apply to more than 50 colleges and 300 college-major options in all the selectivity tiers and the measures discussed above focus on selected tiers (e.g., the ones match students' CEE scores). Furthermore, strategies and preferences are interrelated that students may not change one particular strategy while holding all else unchanged. However, results in this subsection using the constructed college application behavior measures are very consistent with the expectations in designing the project, as well as numerous field observations and feedback not captured in the data.

5.2. Effects on Admission Outcomes

Table 6 presents the results of the ITT effects of the guidebook and workshop interventions on college access and match outcomes. Each column reports coefficient estimates from a separate OLS regression of Equation 1.[41] Both the guidebook and the guidebook-workshop combined interventions substantially improved college admissions. On the extensive margin, Column 1 shows that offering guidebooks or school workshops caused students to be 2 to 3 percentage points more likely to be admitted to a college, although imprecisely estimated due to the small number of clusters. In Table 7, I show that the interventions insignificantly increased college application by about one percentage point. Comparing the two estimates, the interventions have increased the college admission rate conditional on application by about 1 to 2 percentage points. Nevertheless, this increase was not statistically significant.

Results on the intensive margin of college match show that treated students on average were

[41] Results are similar when I do not control for student demographic covariates or control for additional school covariates that are aggregated from student covariates.

admitted to statistically significantly and substantially higher quality colleges. Column 2 of Table 6 shows that students who were offered the "How to Apply to College" guidebook and potentially read it were admitted to a college with a 0.094 *SD* ($p<0.001$) higher quality using the single college match index. Because centralized college admissions solely depend on a student's CEE score and their applications, if students remained unchanged in their college application behaviors, they would have had to score 0.094 *SD* higher on the CEE to be able to get into the same college. This result demonstrates that providing a "college application textbook" generates large improvements in student college access and match during a very short time at a reasonably low cost.

The ITT effects of the guidebook-workshop combined intervention (T2) were very similar. Treated students, on average, were admitted to colleges with a 0.076 *SD* ($p<0.05$) higher college match index. Given the anecdotal evidence, the approximate TOT effects for a student who may have learned from both the guidebook and workshop might be two or three times larger than the guidebook alone. For example, using the Wald estimator, the rescaled TOT effect on college match index is about 0.18 *SD* (ranging from 0.15 to 0.23 SD). The results confirm that informing students about how to make precise predictions of college admission probabilities is effective at helping them improve college match. The intensity added by a school workshop further improved the effectiveness of college-choice advising.

To check the robustness of defining college quality, Column 3 excludes students who were not admitted to college and shows a smaller impact of the interventions, but this result was downward biased. Column 4 shows that treated students were 3 to 4 percentage points less likely to be admitted to undermatched colleges. Columns 5-9 show the itemized results of the college match measures, which are the principle-component factors of the summary index in Column 2. Results show that the improvement in college match is stable using either within-province or national measures.

In Table 7, I explore the treatment effects on additional outcomes. Results are mostly imprecisely estimated due to limited statistical power from the school-level randomization. There is suggestive evidence that the interventions increased college enrollment in the same year by decreasing the probability of repeating the 12th grade for another year.[42] The interventions also increased the share of students who were admitted to match/peer and overmatch/reach colleges.

5.3. Heterogeneity Analysis

Effects on high-achieving students. I found similar intervention effects on high-achieving students. High-achieving students include those who were eligible for admissions at selective (tier 1) colleges based on their CEE scores. As shown in Table B.5, providing both a guidebook and a school workshop to high-achieving students had similarly affected their use of targeting strategies and out-of-province college preferences. The impact of the guidebook-only intervention was smaller and statistically insignificant. Table B.6 repeats the analysis on college admission outcomes. Nearly all high-achieving students, who were also highly motivated for college, applied to and were admitted by a college. Thus, the interventions had a precisely zero effect on applications and admissions.

This finding is different from that in the U.S. For example, the ECO project in Hoxby and Turner (2013) increased high-achieving, low-income students' college admissions by 12%. This is because some high-achieving American students may not apply to any college. Chinese students do apply for college. But they may not know how to select an appropriate set of colleges due to problems of precision predictions of admission opportunities.

As shown in Table B.6, I found clear evidence that both the guidebook and the guidebook-workshop combined interventions statistically significantly improved college match for high-achieving students in both the single-index and itemized measures. Being offered and potentially reading the guidebook increased college match index by 0.058 *SD* ($p<0.05$), holding CEE scores and demographics equal. Being offered guidebook and workshop increased college match index by 0.08 *SD* ($p<0.001$).

Heterogeneity by student characteristics. Figure 2 summarizes the heterogeneous effects on college admissions. I also found similar, consistent heterogeneity in college choice behaviors.

[42] Survey data show that students who choose to repeat are unsatisfied with either their CEE scores or college admission results.

For the guidebook-only intervention, the ITT effects were slightly larger for rural, female, and minority students. For the guidebook-workshop combined intervention, female and non-minority students benefited more. These differences may have resulted from differential take-up between groups. The interventions did not have large impacts on repeaters. Repeaters already had at least one year's experience with college applications. They were more experienced and skilled in searching for and using the relevant information and in strategic decision-making. Figure A.2 shows similar results among high-achieving students. One exception is that high-achieving repeaters also benefited from the workshop, particularly for making precise predictions of college admissions.

5.4. Potential General Equilibrium Effects

One concern regarding the estimated intervention impacts on college admission outcomes is that about one-third of the population of applicants were treated. The competition in college admissions might create a general equilibrium effect. That is, the admission outcomes of the control group were likely to be negatively impacted by the fact that their peers received college application advising and thus improved their admission results. If this is true, the estimated intervention impacts on admissions might be upward biased.

However, this general equilibrium effect is likely to be very small. First, in practice, students have diverse preferences and apply to a large number of colleges and majors. On average, students with the same CEE scores apply to about 300 different colleges and majors. Diverse preferences, meaning that students may apply to different sets of colleges and majors, largely reduce the potential congestion problem that helping some students improve their admission scores dramatically harms other students.

Second, even if there exists congestion that students happen to apply to the same colleges and majors, only a very small number of students might be impacted. The improved college application behaviors of the treated students would only affect those control group students who would have been admitted to overmatched colleges. Table 2 shows that the share of overmatched students was quite small (12% overall, 5% among the top 75% students in the CEE score distribution).

Third, I perform a simulation analysis to provide additional evidence that the potential general equilibrium effect is minimal. I project the changes in admission outcomes using (a) the estimated correlation coefficients between college choice behaviors and college admission outcomes in the untreated sample (reported in Table 3) and (b) the intervention impacts on college choice behaviors (reported in Table 4). This analysis simulates the changes in admission outcomes of the treatment group, holding the correlation between application behaviors and admission outcomes unchanged (i.e., without any general equilibrium effects). The projected increases in the college match index by the guidebook-only and the guidebook-workshop combined interventions are 0.033 *SD* and 0.029 *SD*, which are close to the estimated ITT effects in Column 3 of Table 6 (0.030 *SD* and 0.029 *SD*).

6. Conclusion

In this paper, I studied college choice behaviors and admission outcomes in a centralized college admission system. Using administrative data of college applications and admissions in one of the poorest provinces in China, I documented that the student-college academic undermatch is prevalent in centralized college admissions. The key reason is that students do not make precise predictions of college admission probabilities and use appropriate college application strategies based on such predictions. Using a large-scale randomized experiment, I further show that informing low-income students on how to make precise predictions of college admission probabilities is effective in improving their college access and match. Importantly, except for the location preference targeted by the interventions, the prediction-focused interventions improved college admission without affecting students' college and major preferences.

This paper provides one of the first proof-of-concept evidence on the importance of precise predictions in shaping students' college admission outcomes in centralized systems, which reward strategic and sophisticated college application behaviors. Providing information and guidance is an effective policy intervention for disadvantaged students who often lack such data analytics and decision-making capacities. However, even with the detailed guidance on how to make predictions and how to use the *targeting strategies* described in the paper, the heterogeneity analysis and qualitative evidence in the field suggests that students vary in their abilities to complete the data analytics tasks. While the average treatment effects on college admissions estimated in this paper would likely remain were those treatments to scale up to help every student accurately estimate their admission probabilities, heterogeneous preferences and biased beliefs for colleges and majors may drive additional variations in admission outcomes (Altmejd et al., 2021, Ding et al., 2021; Conlon and Patel, 2022). One possible improvement of the college-choice interventions would be to help optimize predictions and targeting strategies for each student (Ye, 2021; Arteaga et al., 2022). However, individualized advising requires intensive data analytics. Innovative scalable policy solutions are needed to simplify the data analysis process and to increase the scale-up potentials to reduce the inequality in college opportunities.

References

Altmejd, A., Barrios-Fernández, A., Drlje, M., Goodman, J., Hurwitz, M., Kovac, D., Mulhern, C., Neilson, C., and Smith, J. (2021). O brother, where start thou? sibling spillovers on college and major choice in four countries. *The Quarterly Journal of Economics*, 136(3):1831–1886.

Altmejd, A., Bizopoulou, A., Kaila, M., Barrios-Fernández, A., , Neilson, C., Otero, S., and Ye, X. (2022). Inequality in college applications: Evidence from three continents. *Working Paper*.

Arteaga, F., Kapor, A. J., Neilson, C. A., and Zimmerman, S. D. (2022). Smart matching platforms and heterogeneous beliefs in centralized school choice. *The Quarterly Journal of Economics*, 137(3):1791–1848.

Avery, C. and Hoxby, C. M. (2004). Do and should financial aid packages affect students' college choices? In *College choices: The economics of where to go, when to go, and how to pay for it*, pages 239–302. University of Chicago Press.

Bailey, M. J. and Dynarski, S. M. (2011). Inequality in postsecondary education. In Duncan, G. and Murnane, R., editors, *Whither opportunity? Rising inequality, schools, and children's life chances*, pages 117–132. New York: Russell Sage.

Bénabou, R. and Tirole, J. (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, 30(3):141–64.

Benkert, J.-M. and Netzer, N. (2018). Informational requirements of nudging. *Journal of Political Economy*, 126(6):2323–2355.

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

Bettinger, E. P., Long, B. T., Oreopoulos, P., and Sanbonmatsu, L. (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.

Bonilla-Mejía, L., Bottan, N. L., and Ham, A. (2019). Information policies and higher education choices experimental evidence from colombia. *Journal of Behavioral and Experimental Economics*, 83:101468.

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

Bowman, N. A., Kim, S., Ingleby, L., Ford, D. C., and Sibaouih, C. (2018). Improving college access at low-income high schools? the impact of gear up iowa on postsecondary enrollment and persistence. *Educational Evaluation and Policy Analysis*, page 0162373718778133.

Campbell, S., Macmillan, L., Murphy, R., and Wyness, G. (2022). Matching in the dark? inequalities in student to degree match. *Journal of Labor Economics*, 40(4):807–850.

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

Carruthers, C. K. and Fox, W. F. (2016). Aid for all: College coaching, financial aid, and post-secondary persistence in tennessee. *Economics of Education review*, 51:97–112.

Castleman, B. (2017). Behavioral insights for federal higher education policy. *Urban Institute*.

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

Castleman, B. L., Owen, L., and Page, L. C. (2015a). 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, B. L., Schwartz, S., and Baum, S. (2015b). *Decision making for student success: Behavioral insights to improve college access and persistence*. Routledge.

Chade, H., Lewis, G., and Smith, L. (2014). Student portfolios and the college admissions problem. *Review of Economic Studies*, 81(3):971–1002.

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

Conlon, J. J. and Patel, D. (2022). What jobs come to mind? stereotypes about fields of study. *Working Paper*.

Corcoran, S. P., Jennings, J. L., Cohodes, S. R., and Sattin-Bajaj, C. (2018). Leveling the playing field for high school choice: Results from a field experiment of informational interventions. *NBER Working Paper No. 24471*.

Damgaard, M. T. and Nielsen, H. S. (2018). Nudging in education. *Economics of Education Review*, 64:313–342.

DesJardins, S. L., Ahlburg, D. A., and McCall, B. P. (2006). An integrated model of application, admission, enrollment, and financial aid. *The Journal of Higher Education*, 77(3):381–429.

Dillon, E. W. and Smith, J. A. (2017). The determinants of mismatch between students and colleges. *Journal of Labor Economics*, 35(1):45–66.

Dillon, E. W. and Smith, J. A. (2020). The consequences of academic match between students and colleges. *Journal of Human Resources*, 55(3):767–808.

Ding, Y., Li, W., Li, X., Wu, Y., Yang, J., and Ye, X. (2021). Heterogeneous major preferences for extrinsic incentives: The effects of wage information on the gender gap in stem major choice. *Research in Higher Education*, 62(8):1113–1145.

Dinkelman, T. and Martínez A, C. (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, S., Libassi, C., Michelmore, K., and Owen, S. (2021). Closing the gap: The effect of reducing complexity and uncertainty in college pricing on the choices of low-income students. *American Economic Review*, 111(6):1721–56.

Dynarski, S., Nurshatayeva, A., Page, L. C., and Scott-Clayton, J. (2022). Addressing non-financial barriers to college access and success: Evidence and policy implications. *NBER Working Paper No.30054*.

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

Evans, B. J. and Boatman, A. (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*, pages 1–33.

French, R. and Oreopoulos, P. (2017). Behavioral barriers transitioning to college. *Labour Economics*, 47:48–63.

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

Gurantz, O., Pender, M., Mabel, Z., Larson, C., and Bettinger, E. (2019). Virtual advising for high-achieving high school students. *EdWorkingPaper No. 19-126*.

Hastings, J., Neilson, C. A., and Zimmerman, S. D. (2018). The effects of earnings disclosure on college enrollment decisions. *NBER Working Paper No. 21300*.

Herber, S. P. (2018). The role of information in the application for highly selective scholarships: Evidence from a randomized field experiment. *Economics of Education Review*, 62:287–301.

Heß, S. (2017). Randomization inference with stata: A guide and software. *Stata Journal*, 17(3):630–651.

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

Holsinger, D. B. and Jacob, W. J. (2009). *Inequality in education: Comparative and international perspectives*, volume 24. Springer Science & Business Media.

Hossler, D. and Gallagher, K. S. (1987). Studying student college choice: A three-phase model and the implications for policymakers. *College and University*, 62(3):207–21.

House, E. (2017). Finding the best fit: Exploring postsecondary undermatch in tennessee. *Ph.D. Dissertation at the University of Michigan*.

Howell, J. S. and Pender, M. (2016). The costs and benefits of enrolling in an academically matched college. *Economics of Education Review*, 51:152–168.

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

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

Hoxby, C. M. and Turner, S. (2013). Expanding college opportunities for high-achieving, low-income students. *SIEPR Discussion Paper No. 12-014*.

Hurwitz, M., Mbekeani, P. P., Nipson, M. M., and Page, L. C. (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.

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

Jacob, B., McCall, B., and Stange, K. (2018). College as country club: Do colleges cater to students' preferences for consumption? *Journal of Labor Economics*, 36(2):309–348.

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

Kang, C. and Torres, D. G. (2018). College undermatching, degree attainment, and minority students. *Paper presented at the 2018 AERA Annual Meeting*.

Kang, L., Ye, X., and Ding, Y. (2020). Traditional dietary habits and interprovincial migration in college choice: Evidence from ningxia in china. *ECNU Review of Education*, page 2096531119889875.

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

Larroucau, T., Martínez, M., Neilson, C., and Rios, I. (2021). Application mistakes and information frictions in college admissions.

Lavecchia, A. M., Liu, H., and Oreopoulos, P. (2016). Behavioral economics of education: Progress and possibilities. In *Handbook of the Economics of Education*, volume 5, pages 1–74. Elsevier.

Li, H., Loyalka, P., Rozelle, S., Wu, B., and Xie, J. (2015). Unequal access to college in china: How far have poor, rural students been left behind? *The China Quarterly*, 221:185–207.

Li, X., Ye, X., and Wu, X. (2022). How college-student academic undermatch affects students: Quasi-experimental evidence on multifaceted student outcomes. *Working Paper*.

List, J. A., Shaikh, A. M., and Xu, Y. (2016). Multiple hypothesis testing in experimental economics. *NBER Working Paper No. 21875*.

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

Loyalka, P., Wu, B., and Ye, X. (2017). Out of the gauntlet and into the fire: How college choices in centralized admissions systems disadvantage disadvantaged students. *Working paper*.

Manski, C. F. and Wise, D. A. (1983). *College choice in America*. Harvard University Press.

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

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

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

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

Ovink, S., Kalogrides, D., Nanney, M., and Delaney, P. (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, L. C., Kehoe, S. S., Castleman, B. L., and Sahadewo, G. A. (2017). More than dollars for scholars: The impact of the dell scholars program on college access, persistence and degree attainment. *Journal of Human Resources*, pages 0516–7935r1.

Page, L. C. and Scott-Clayton, J. (2016). Improving college access in the united states: Barriers and policy responses. *Economics of Education Review*, 51:4–22.

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

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

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

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

Peter, F., Spiess, C. K., and Zambre, V. (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.*

Peter, F. H. and Zambre, V. (2017). Intended college enrollment and educational inequality: Do students lack information? *Economics of Education Review*, 60:125–141.

Prakhov, I. and Sergienko, D. (2017). Matching between students and universities: What are the sources of inequalities of access to higher education? *Higher School of Economics Research Paper No. WP BRP 45/EDU/2017*.

Roderick, M., Nagaoka, J., Coca, V., Moeller, E., Roddie, K., Gilliam, J., and Patton, D. (2008). From high school to the future: Potholes on the road to college. *Chicago, IL: Consortium on Chicago School Research at the University of Chicago*.

Rodriguez, A. (2015). Tradeoffs and limitations: Understanding the estimation of college undermatch. *Research in Higher Education*, 56(6):566–594.

Smith, J., Pender, M., and Howell, J. (2013). The full extent of student-college academic undermatch. *Economics of Education Review*, 32:247–261.

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

Wang, A., Wang, S., and Ye, X. (2022). Cognitive distortions in complex decisions: Evidence from centralized college admission. *Working paper*.

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, X. (2021). Personalized advising for college match: Experimental evidence on the use of human expertise and machine learning to improve college choice.

Figures

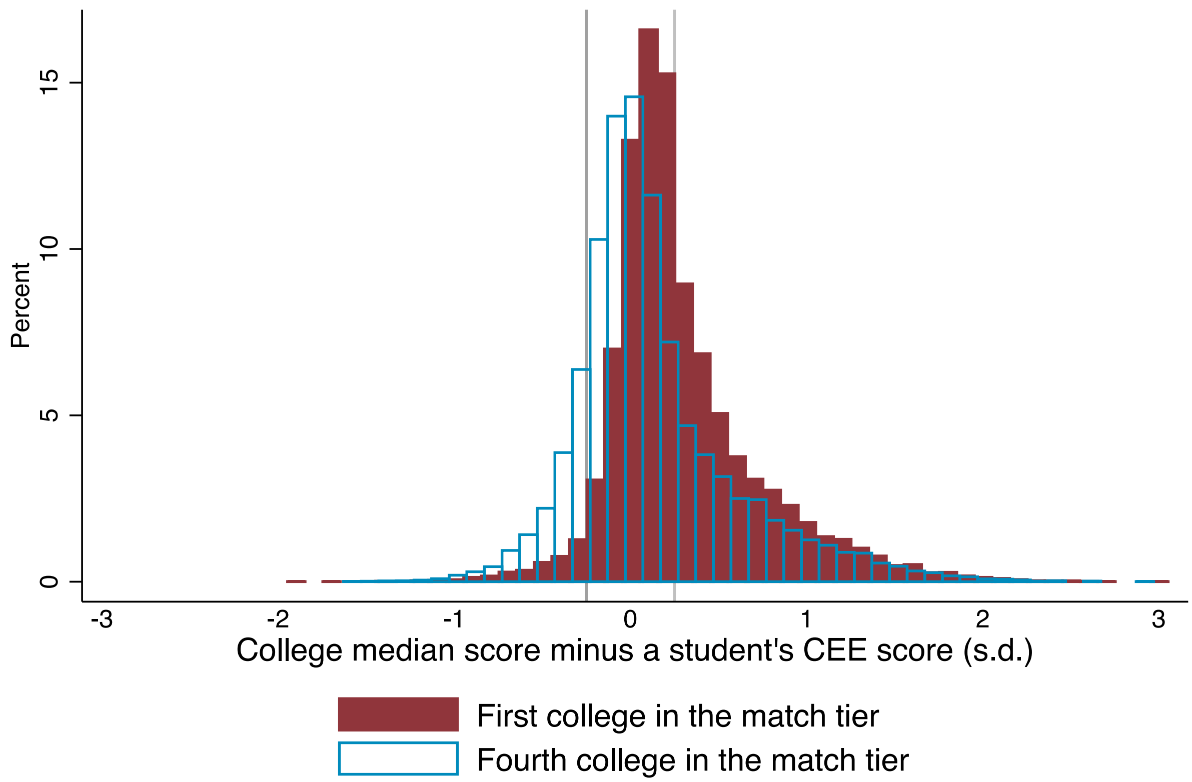
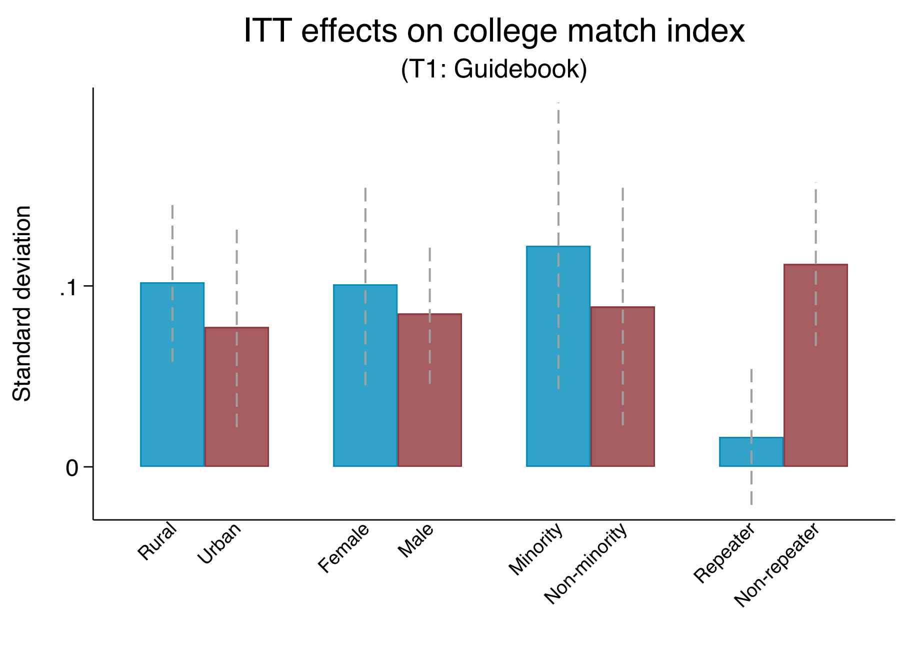
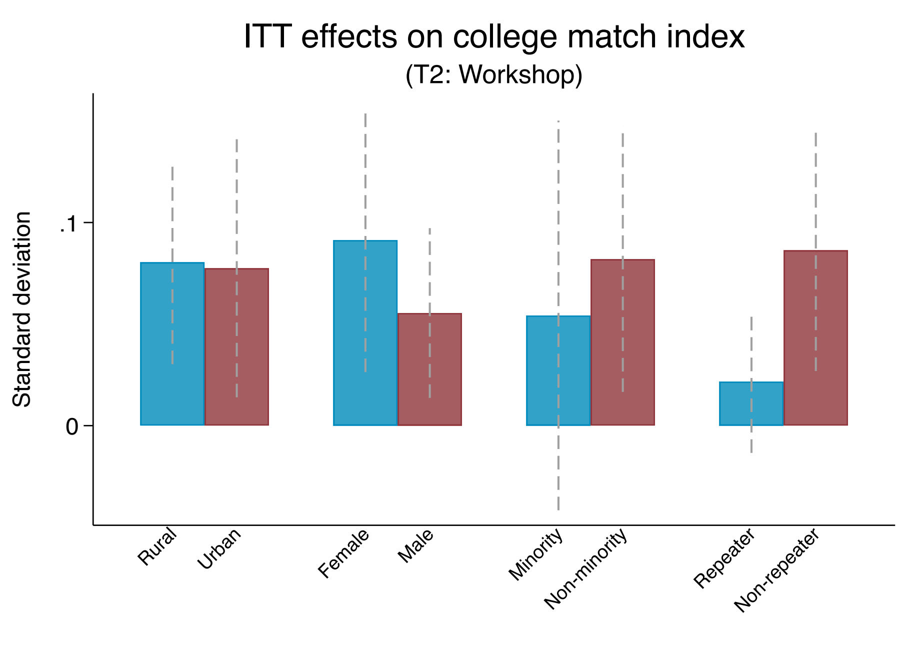


Figure 1: Distribution of the distance of college median score and a student's own score

Notes: This figure shows the distribution of students' applications using the full application data.The X axis shows the distance of college median score and a student's own CEE score. We separately present the distributions for students' first choice and fourth (last) choice in the match tier. **The match tier** indicates the highest possible selectivity tier that one student qualifies for based on her CEE score, which should be her primarily targeting tier. Two vertical gray lines indicate the boundary of match range (0.25 s.d. from zero).



(a) T1: Guidebook



(b) T2: Workshop (based on guidebook)

Figure 2: Heterogeneity in the ITT effects

Notes: This figure plots heterogeneous ITT effects of the interventions on college median score from the OLS regression Equation 1 using each subsample (e.g., rural students vs. urban students) separately. Dashed gray lines indicate 95% confidence intervals.

Tables

Table 1: Experimental design

Sample	All 32,834 students in 31 public high schools in three cities				
Groups	Intervention	Randomization unit (schools)	Students	Take-up	Estimation
Control	No	12	11,408		
Treatment 1	Guidebook	12	12,823	Nearly 100%	ITT
Treatment 2	School workshop	7	8,603	42% (29%-56% by school)	ITT

Notes: This table shows the experimental design of the *Bright Future of China Project* in Ningxia in 2016. The primary randomization is between the control group and the first two treatment groups. Take-up rates for guidebook and school workshop in 2016 are from anecdotal evidence (school survey and field observations).

Table 2: Extent of academic undermatch and overmatch

CEE quartiles (access to)	College quality quartiles (admitted) (N=31,777) 1st Quartile (Highest)	2nd Quartile	3rd Quartile	4the Quartile (Lowest)	No college	Percent Undermatch (0.25 s.d.)	(5 pctl)	Percent Overmatch (0.25 s.d.)
1st Quartile (Highest)	7,450 (90.3)	740 (9.0)	13 (0.2)	1 (0.0)	50 (0.6)	15.0	20.3	4.1
2nd Quartile	1,031 (12.5)	4,837 (58.6)	618 (7.5)	179 (2.2)	1,593 (19.3)	29.1	35.1	7.8
3rd Quartile	13 (0.2)	736 (9.4)	3,714 (47.5)	1,379 (17.6)	1,977 (25.3)	45.9	52.7	4.3
4th Quartile (Lowest)	8 (0.1)	64 (0.9)	1,038 (13.9)	4,955 (66.6)	1,381 (18.6)	25.2	26.1	33.3
Total						28.6	33.5	12.0

Notes: This table reports the joint distribution of students' College Entrance Exam (CEE) score and their admitted colleges' quality (measured by college median CEE score), using the universe of the untreated sample (including both the control group of the randomization sample and those not in the randomization sample) of the 2016 cohort of high school graduates in Ningxia, China. Each cell contains the number of students and the row percentage (in parentheses). The last three columns report the undermatch and overmatch percents by student CEE score quartile, using 5 percentile and 0.25 standard deviation as cutoffs, respectively. **Undermatch** is when a student's own CEE score is 0.25 standard deviation (or 5 rank percentile) higher than her admitted college's median CEE score, or a student was not admitted to any colleges. **Overmatch** is when a student's own CEE score is 0.25 standard deviation lower than her admitted college's median CEE score.

Table 3: College application behaviors and admission outcomes

		Outcome: Index of college match			
		(1)	(2)	(3)	(4)
Strategy	General advice	0.030***	0.006	0.006	0.004
		(0.007)	(0.007)	(0.007)	(0.007)
Strategy	Targeting		0.218***	0.218***	0.217***
			(0.011)	(0.011)	(0.011)
Strategy	Special programs			-0.003	-0.007
				(0.005)	(0.005)
Preference	Tuition & quota				0.057***
					(0.007)
Preference	Location				0.079***
					(0.004)
Preference	Major				-0.006
					(0.004)
Observations		28,806	28,806	28,806	28,806
R-squared		0.666	0.706	0.706	0.710

Notes: This table reports the OLS regression results of the correlations between college application behaviors and college match index, using data from those who submitted college applications in the untreated sample in 2016. Because these measures increase with both matched and overmatched admissions results, the label of "college match measure" indicates the comparison between match/overmatch and undermatch. Application behaviors are constructed using the full applications data, as described in Appendix Subsection C.1. **General advice** strategies include number of applied colleges, percent of applied majors, and percent of flexible major assignment. **Targeting** strategies include equating CEE scores (estimated by the gap between students' CEE scores and their applied colleges' admissions scores), applying to colleges in the match tier, applying to colleges without historical admissions data, ranking colleges in a descending order, and applying to a mix of reach, safety, and match colleges. **Special programs** strategies include indicators of applying to a set of affirmative action and special programs. **Tuition & quota** preferences include measures of tuition and quota of the college that a student applied to. **Location** preferences describe the geographic locations of colleges that a student applied to. **Major** preferences include a set of indicators of majors that a student applied to. All regressions include a student's CEE score, demographic covariates, and high school fixed effects. Standard errors in parentheses are clustered at high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 4: ITT effects on college choice behaviors: Principal-component factors

	Index	Strategy			Preference		
		General	Targeting	Special programs	Tuition & quota	Location	Major
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Control mean	0.171	0.066	0.087	0.299	-0.068	0.130	0.035
Control sd	[0.977]	[0.952]	[1.034]	[1.020]	[1.035]	[0.998]	[1.049]
T1 (guidebook)	0.091	0.071	0.107**	-0.099	-0.024	0.124	0.009
	(0.195)	(0.317)	(0.020)	(0.145)	(0.766)	(0.113)	(0.870)
T2 (workshop)	0.167**	0.040	0.091*	0.076	-0.114	0.208**	0.063
	(0.036)	(0.615)	(0.082)	(0.369)	(0.185)	(0.019)	(0.328)
N	29,591	29,591	29,591	29,591	29,591	29,591	29,591

Notes: This table reports the OLS regression (Equation 1) results of the ITT effects of the guidebook and workshop interventions in 2016 on college choice behaviors. Sample includes all the students in the randomization sample and submitted their college applications. We use principal component factor analysis to create a single index for each strategy and preference group, and an index for all (in column 1). Strategies and preferences are constructed using college application data, as described in Appendix Subsection C.1. All regressions control for student-level covariates (CEE score and demographics) and strata fixed effects. After rotation in the principal component analysis, the regression coefficients used to estimate individual factor index scores (reported in column 1) for each of these six application behavior indices (reported in columns 2-7) are 0.224, 0.307, 0.183, -0.335, 0.393, and 0.174. The scoring coefficients sum up to 1. Minor differences exist due to rounding. Randomization inference p-values from 1,000 permutations are in parentheses (clustered at high school level). * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 5: ITT effects on college choice behaviors: Itemized results

	General advice			Targeting				
	# College	% major	% flexible	Estimated gap	No match tier	Descending	Targeting	Missing prior data
		(%)	(%)	(=1)	(=1)	(=1)	(=1)	(=1)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Control mean	7.898	70.723	70.196	0.384	0.238	0.338	0.366	0.018
Control sd	[4.490]	[22.626]	[36.366]	[0.486]	[0.426]	[0.473]	[0.482]	[0.133]
T1 (guidebook)	-0.655*	1.078	2.778	0.020**	-0.033	0.044**	0.042**	0.004
	(0.053)	(0.347)	(0.371)	(0.044)	(0.179)	(0.011)	(0.040)	(0.131)
T2 (workshop)	0.116	1.065	0.756	0.015	-0.034	0.035*	0.034*	0.004
	(0.776)	(0.387)	(0.849)	(0.182)	(0.192)	(0.058)	(0.096)	(0.190)

	Special programs			Tuition and quota	
	AA	Early	Teachers	Tuition	Quota
	(%)	(%)	(%)	(in 1000s)	
	(1)	(2)	(3)	(4)	(5)
Control mean	0.275	0.197	3.788	6.233	655.006
Control sd	[0.446]	[0.398]	[9.461]	[3.125]	[566.357]
T1 (guidebook)	-0.031*	0.001	1.223	-0.187	-58.934
	(0.077)	(0.958)	(0.111)	(0.326)	(0.226)
T2 (workshop)	-0.004	0.029	-0.963	0.058	-86.557
	(0.862)	(0.235)	(0.363)	(0.777)	(0.132)

	Location			Major				
	Out of province	Developed regions	Neighborhood	Economics	Agriculture	CS	International	Medical
	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Control mean	40.258	7.806	51.936	23.341	1.246	2.947	1.917	11.656
Control sd	[31.967]	[14.795]	[35.252]	[20.877]	[4.440]	[5.710]	[4.259]	[20.343]
T1 (guidebook)	3.195	1.151	-4.346	-0.458	-0.017	0.361	-0.077	-1.701
	(0.299)	(0.196)	(-0.167)	(0.627)	(0.898)	(0.145)	(0.609)	(0.124)
T2 (workshop)	5.209	1.977**	-7.186**	1.201	-0.141	0.516*	0.124	-1.078
	(0.128)	(0.029)	(0.047)	(0.360)	(0.404)	(0.079)	(0.470)	(0.437)

Notes: This table reports the OLS regression (Equation 1) results of the ITT effects of the guidebook and workshop interventions in 2016 on college choice behaviors (detailed items). Strategies and preferences are constructed using college application data, as described in Appendix Subsection C.1. Sample includes all the students in the randomization sample and submitted their college applications. All regressions control for student-level covariates (CEE score and demographics) and strata fixed effects. Randomization inference p-values from 1,000 permutations are in parentheses (clustered at high school level). * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 6: ITT effects on college access and match outcomes

	Main outcomes				Outcomes in college match index (column 2)				
	Admission to a college (=1)	College match index (s.d.)	College match index* (s.d.)	Undermatch (=1)	College median score (s.d.)	College mean score (s.d.)	College minimum score (s.d.)	College quality measure (s.d.)	College ranking (pctl)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Control mean	0.846	0.182	0.263	0.296	0.052	0.068	-0.688	-0.151	55.835
Control s.d.	[0.361]	[0.991]	[0.902]	[0.457]	[1.142]	[1.115]	[1.333]	[1.872]	[34.365]
				A. Without school covariates					
T1 (guidebook)	0.032*	0.094***	0.030*	-0.040**	0.089**	0.083**	0.171***	0.181**	2.456**
	(0.076)	(0.009)	(0.075)	(0.033)	(0.017)	(0.020)	(0.000)	(0.023)	(0.029)
T2 (workshop)	0.024	0.076**	0.029	-0.026	0.071*	0.067*	0.118**	0.156*	2.324**
	(0.276)	(0.044)	(0.134)	-0.231	(0.088)	(0.090)	(0.040)	(0.075)	(0.045)
				B. With school covariates					
T1 (guidebook)	0.040**	0.114***	0.034**	-0.045**	0.115***	0.107***	0.187***	0.229***	2.997**
	(0.033)	(0.005)	(0.045)	(0.028)	(0.007)	(0.007)	(0.002)	(0.009)	(0.012)
T2 (workshop)	0.033	0.085**	0.020	-0.027	0.088*	0.083*	0.112*	0.191**	2.381*
	(0.128)	(0.048)	(0.359)	(0.227)	(0.059)	(0.058)	(0.072)	(0.046)	(0.077)
N	32,834	32,834	27,657	32,834	32,834	32,834	27,657	32,834	32.834

Notes: This table reports the OLS regression (Equation 1) results of the ITT effects of the guidebook and workshop interventions in 2016 on a family of college access and match outcomes. **Admission to a college** denotes whether a student was admitted to college. **College match index** measures college match, using principal component factor analysis based the five continuous outcomes in columns (5)-(9). After rotation, the regression coefficients used to estimate individual factor index scores for each of these five measures are 0.218, 0.218, 0.192, 0.210, and 0.211. The scoring coefficients sum up to 1. Minor differences exist due to rounding. **College match index*** excludes students who were not admitted to college. **Undermatch** is when a student's own CEE score is 0.25 standard deviation higher than here admitted college's median CEE score, or a student was not admitted to any colleges. **College median/mean/minimum scores** are constructed using all the admissions data in Ningxia in 2016. **College quality measure (standardized)** describes college quality using national data on college (admissions scores, inputs and employment data) from 1996-2017, and **College ranking** is the corresponding ranking percentile. All regressions control for student-level covariates (CEE score and demographics) and strata fixed effects. Randomization inference p-values from 1,000 permutations are in parentheses (clustered at high school level). * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 7: ITT effects on additional college access and match outcomes

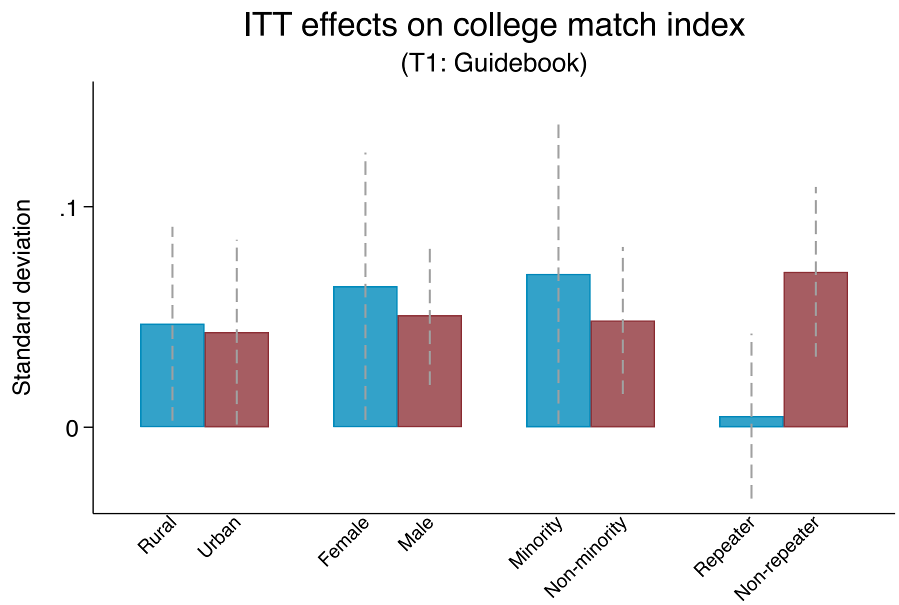
	Application (=1) (1)	Enrollment in 2016 (=1) (2)	Repeating in 2017 (=1) (3)	Match (=1) (4)	Overmatch (=1) (5)
Control mean	0.914	0.777	0.206	0.608	0.096
Control s.d.	[0.280]	[0.416]	[0.405]	[0.488]	[0.294]
T1 (guidebook)	0.012	0.022	-0.028	0.035*	0.005
	(0.467)	(0.358)	(0.287)	(0.055)	(0.619)
T2 (workshop)	0.012	0.030	-0.052	0.016	0.010
	(0.573)	(0.304)	(0.104)	(0.447)	(0.398)
N	32,834	32,834	32,834	32,834	32,834

Notes: This table reports the OLS regression (Equation 1) results of the ITT effects of the guidebook and workshop interventions in 2016 on additional college access and match outcomes. **Enrollment in 2016** denotes students who received college admissions and did not repeat in 2017 (we do not have data from colleges about their actual enrollment status). **Repeating in 2017** denotes students who took CEE in 2016 and in 2017. **Match** indicates that a student's admitted college median score is within 0.25 s.d. radius of her own CEE score. All regressions control for student-level covariates (CEE score and demographics) and strata fixed effects. Randomization inference p-values from 1,000 permutations are in parentheses (clustered at high school level). * significant at 10%, ** significant at 5%, *** significant at 1%.

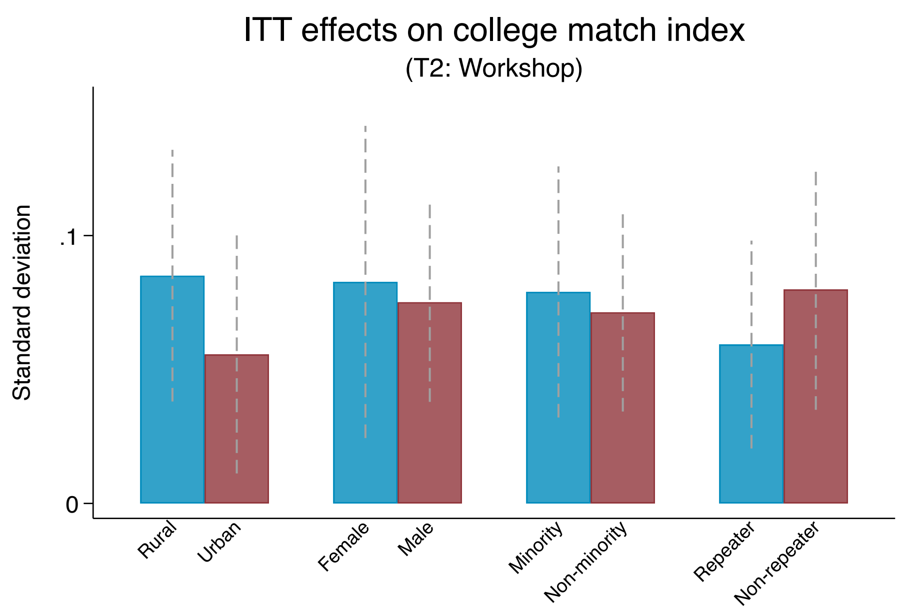
A. Appendix Figures

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.



(a) T1: Guidebook



(b) T2: Workshop

Figure A.2: Heterogeneity in the ITT effects: High achieving students

Notes: This figure plots heterogeneous ITT effects among high-achieving students of the interventions on college median score from the OLS regression Equation 1, but with each subsample (e.g., rural students vs. urban students) separately. High-achieving students include those who were eligible for admissions at selective (tier 1) colleges. Dashed gray lines indicate 95% confidence intervals.

B. Appendix Tables

Table B.1: Measures of undermatch: Varying thresholds

	% undermatch							
	Including not admitted students				Excluding not admitted students			
CEE quartiles	0.05 s.d.	0.15 s.d.	0.25 s.d.	0.35 s.d.	0.05 s.d.	0.15 s.d.	0.25 s.d.	0.35 s.d.
1st Quartile (Highest)	45.1	25.7	15.0	9.4	44.7	25.3	14.5	8.8
2nd Quartile	45.9	33.5	29.1	27.4	33.0	17.6	12.1	10.1
3rd Quartile	63.5	53.1	45.9	41.8	51.1	37.2	27.6	22.1
4th Quartile (Lowest)	35.7	29.4	25.2	22.5	21.1	13.3	8.1	4.8
Total	47.6	35.3	28.6	25.1	37.9	23.3	15.3	11.1

Notes: This table shows the distribution of undermatch in different student CEE score quartiles along with varying thresholds. Even using a very conservative threshold (0.35 standard deviation above the college median CEE score) to define undermatch and focusing on the selected sample of students who were already admitted to college, there is still a substantial proportion of students were admitted to academically undermatched colleges.

Table B.2: Sample description

	All	Not in RCT sample	RCT sample	Control	T1	T2
	(1)	(2)	(3)	(4)	(5)	(6)
Schools	60	29	31	12	12	7
Students	56,172	23,338	32,834	11,408	12,823	8,603
Rural	0.59	0.55	0.62	0.56	0.71	0.57
Female	0.55	0.55	0.54	0.53	0.56	0.54
Minority	0.31	0.34	0.28	0.38	0.24	0.21
Age (>=18)	0.87	0.86	0.87	0.84	0.90	0.86
STEM	0.67	0.65	0.68	0.70	0.66	0.69
Repeater	0.19	0.18	0.20	0.15	0.25	0.19
CEE score	0.09	0.03	0.14	0.36	-0.07	0.15
Admitted	0.84	0.84	0.84	0.85	0.84	0.84
College median score	-0.17	-0.21	-0.15	0.05	-0.34	-0.13

Notes: This table describes the sample in the 2016 program. Randomization is at school-level within strata. The descriptive statistics do not account for between-strata differences.

Table B.3: Balance checks

	All students			High achieving students		
	Control	T1	T2	Control	T1	T2
	(1)	(2)	(3)	(4)	(5)	(6)
A. Student-level results using student data in 2016						
Rural	0.556	-0.001	-0.133	0.650	-0.218	-0.294
	[0.497]	(0.997)	(0.282)	[0.477]	(0.268)	(0.101)
Female	0.526	0.017	-0.003	0.500	0.025	0.004
	[0.499]	(0.308)	(0.887)	[0.500]	(0.294)	(0.922)
Minority	0.384	-0.131	-0.161	0.452	-0.208	-0.172
	[0.486]	(0.212)	(0.186)	[0.498]	(0.045)	(0.119)
Age	0.842	0.029	-0.012	0.819	-0.020	-0.062
	[0.365]	(0.323)	(0.711)	[0.385]	(0.718)	(0.276)
STEM	0.697	-0.015	0.021	0.811	-0.053	-0.002
	[0.459]	(0.684)	(0.631)	[0.392]	(0.163)	(0.969)
Repeater	0.146	0.034	-0.028	0.139	-0.011	-0.133
	[0.353]	(0.368)	(0.623)	[0.346]	(0.874)	(0.289)
CEE score	0.364	-0.120	0.141	1.237	0.082	0.159
	[0.852]	(0.571)	(0.581)	[0.402]	(0.390)	(0.205)
B. School-level results (unweighted) using student data in 2016						
Rural	0.556	0.121	0.002	0.650	0.113	-0.013
	[0.311]	(0.257)	(0.990)	[0.305]	(0.352)	(0.930)
Female	0.526	0.015	0.032*	0.500	0.037	-0.002
	[0.034]	(0.454)	(0.062)	[0.036]	(0.204)	(0.925)
Minority	0.384	-0.044	-0.083	0.452	-0.038	-0.040
	[0.159]	(0.602)	(0.422)	[0.173]	(0.673)	(0.759)
Age	0.842	0.039	0.000	0.819	0.021	-0.042
	[0.053]	(0.125)	(0.994)	[0.062]	(0.647)	(0.431)
STEM	0.697	-0.002	0.033	0.811	-0.525	0.008
	[0.082]	(0.960)	(0.554)	[0.057]	(0.278)	(0.874)
Repeater	0.146	0.016	-0.022	0.139	0.043	-0.130
	[0.056]	(0.616)	(0.542)	[0.134]	(0.668)	(0.271)
CEE score	0.364	-0.087	0.173	1.237	0.021	0.078
	[0.408]	(0.681)	(0.507)	[0.099]	(0.696)	(0.200)
C. School data in 2013						
Students	3,016.1	-340.8	332.6			
	[1,953.2]	(0.653)	(0.662)			
Full-time teachers	204.5	-6.3	64.1			
	[144.2]	(0.891)	(0.186)			
Part-time teachers	11.3	-3.9	-7.9			
	[15.7]	(0.548)	(0.245)			
Buildings	13.9	-2.9	-3.7			
	[7.7]	(0.461)	(0.370)			
Assets (in 1000)	24.6	-1.8	-5.0			
	[21.1]	(0.774)	(0.582)			
Books	5.2	2.9	5.6			
	[7.9]	(0.633)	(0.247)			
Total revenue	12,170.8	632.0	1,668.7			
	[3,754.7]	(0.740)	(0.356)			
Fiscal revenue	8,578.6	-304.5	546.7			
	[2,318.8]	(0.798)	(0.688)			
Tuitions	1,143.7	-326.3	-545.0			
	[756.8]	(0.371)	(0.159)			
Total spending	12,686.5	711.5	2,237.1			
	[3,868.8]	(0.682)	(0.289)			
Salary spending	2,035.9	-334.6	-61.0			
	[895.5]	(0.206)	(0.788)			
Operation spending	2,205.0	244.1	276.7			
	[1,103.9]	(0.574)	(0.561)			

Notes: This table reports the balance checks results using student/school-level observations of student data in 2016, and school finance data in 2013. The latter was used for randomization and initial balance checks. Random inference (and its p-value, reported in parentheses) is from 1,000 times permutations. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table B.4: Balance checks: Prediction of treatment status using student-level covariates

	All students		High achieving students	
	T1	T2	T1	T2
	(1)	(2)	(3)	(4)
Rural	-0.007	-0.019	-0.130	-0.007
	(0.077)	(0.059)	(0.130)	(0.041)
Female	0.010	0.001	0.004	0.006
	(0.007)	(0.011)	(0.011)	(0.010)
Minority	-0.112*	-0.106	-0.122*	-0.057
	(0.060)	(0.073)	(0.062)	(0.044)
Age	0.030	0.028	0.020	0.023
	(0.019)	(0.018)	(0.018)	(0.014)
STEM	-0.009	0.025	-0.021	0.057
	(0.023)	(0.039)	(0.043)	(0.043)
Repeater	0.027	-0.033	-0.027	-0.052
	(0.037)	(0.023)	(0.032)	(0.062)
CEE score	-0.015	0.023	0.057	0.069
	(0.027)	(0.045)	(0.038)	(0.042)
2.STRATA	0.366	0.546**	0.200	0.407
	(0.232)	(0.238)	(0.269)	(0.253)
3.STRATA	0.658***	0.709***	0.692***	0.885***
	(0.223)	(0.227)	(0.214)	(0.113)
4.STRATA	0.667***	0.696**	0.707***	0.771***
	(0.206)	(0.254)	(0.186)	(0.218)
Constant	0.130	0.002	0.185	-0.117
	(0.138)	(0.027)	(0.235)	(0.075)
F test	1.215	1.946	0.745	1.094
(P value)	0.335	0.121	0.637	0.407
Observations	24,231	20,011	5,831	5,738
R-squared	0.395	0.426	0.478	0.680

Notes: This table reports the balance checks results from separate OLS regressions that predict the treatment status using student-level data in 2016. Each column is from a separate regression. Strata fixed effects are included. Joint F test results are reported at the bottom of the table. Standard errors in parentheses are clustered at high schools. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table B.5: ITT effects on college choice behaviors for high achieving students: Principal-component factors

	Index	Strategy			Preference		
		General	Targeting	Special programs	Tuition & quota	Location	Major
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Control mean	0.701	0.174	0.572	0.803	-0.009	0.677	-0.061
Control sd	[0.878]	[0.887]	[0.881]	[1.030]	[0.684]	[1.009]	[1.065]
T1 (guidebook)	0.093	0.084	0.074	-0.051	-0.118	0.144	0.044
	(0.273)	(0.358)	(0.137)	(0.653)	(0.140)	(0.140)	(0.293)
T2 (workshop)	0.229**	0.091	0.134**	-0.003	-0.219**	0.309***	0.074
	(0.016)	(0.326)	(0.034)	(0.969)	(0.014)	(0.010)	(0.164)
N	7,973	7,973	7,973	7,973	7,973	7,973	7,973

Notes: This table reports the OLS regression (Equation 1) results of the ITT effects of the guidebook and workshop interventions in 2016 on college choice behaviors. Strategies and preferences are constructed using college application data, as described in Appendix Subsection C.1. Sample includes high-achieving students in the randomization sample and submitted their college applications. High-achieving students include those who were eligible for admissions at selective (tier 1) colleges. We use principal component factor analysis to create a single index for each strategy and preference group, and an index for all (in column 1). All regressions control for student-level covariates (CEE score and demographics) and strata fixed effects. Randomization inference p-values from 1,000 permutations are in parentheses (clustered at high school level). * significant at 10%, ** significant at 5%, *** significant at 1%.

Table B.6: ITT effects on college access and match outcomes for high achieving students

A. Main outcomes

	Admission (=1) (1)	Index (s.d.) (2)	Index* (s.d.) (3)	Undermatch (=1) (4)
Control mean	0.998	1.122	1.075	0.192
Control s.d.	[0.044]	[0.361]	[0.428]	[0.394]
T1 (guidebook)	0.001	0.058**	0.066**	-0.038
	(0.513)	(0.030)	(0.036)	(0.178)
T2 (workshop)	0.001	0.080***	0.092***	-0.066**
	(0.078)	(0.010)	(0.007)	(0.028)
N	7,977	7,977	7,961	7,977

B. Outcomes in Index (in column 2)

	College median (s.d.) (5)	College mean (s.d.) (6)	College min (s.d.) (7)	Quality (s.d.) (8)	Ranking (pctl) (9)
Control mean	1.130	1.130	0.285	1.447	89.902
Control s.d.	[0.449]	[0.436]	[1.202]	[0.490]	[7.184]
T1 (guidebook)	0.041*	0.042*	0.216***	0.050**	0.804**
	(0.067)	(0.058)	(0.005)	(0.041)	(0.038)
T2 (workshop)	0.056**	0.054**	0.298**	0.065**	1.215**
	(0.026)	(0..026)	(0.012)	(0.021)	(0.018)

C. Other outcomes

	Application (=1) (10)	Enrollment in 2016 (=1) (11)	Repeating in 2017 (=1) (12)	Match (=1) (13)	Overmatch (=1) (14)
Control mean	1.000	0.978	0.022	0.768	0.040
Control s.d.	[0.017]	[0.148]	[0.148]	[0.422]	[0.196]
T1 (guidebook)	-0.000	0.011*	-0.011*	0.027	0.012
	(0.937)	(0.053)	(0.056)	(0.303)	(0.160)
T2 (workshop)	-0.001	0.003	-0.003	0.057*	0.009
	(0.677)	(0.653)	(0.639)	(0.056)	(0.317)

Notes: This table reports the OLS regression (Equation 1) results of the ITT effects of the guidebook and workshop interventions in 2016 on a family of college access and match outcomes for high-achieving students. High-achieving students include those who were eligible for admissions at selective (tier 1) colleges. Outcomes are the same as described previously. All regressions control for student-level covariates (CEE score and demographics) and strata fixed effects. Randomization inference p-values from 1,000 permutations are in parentheses (clustered at high school level). * significant at 10%, ** significant at 5%, *** significant at 1%.

C. Additional Descriptions

C.1. Measuring College Application Behaviors and Their Correlations with College Admissions

C.1.1 Tier-Specific College Applications in Chinese Centralized Admissions

As introduced in Subsection 2.1, college applications and admissions in China proceed by institutional selectivity tiers within province-track. 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 colleges 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 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).[43]

[43]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 application form in Ningxia in 2016 (Simplified)

ID: Name: Track:

Tier	**No.**	**College**	**Major**						**Flexible assignment?**
			1	2	3	4	5	6	
Tier 1 - Early Admissions	1								
	2								
Tier 2 - Early Admissions	1								
	1								
Tier 1 - National Affirmative Action (Rural)	A								
	B								
	C								
Tier 1	A								
	B								
	C								
	D								
Tier 1 - Provincial Affirmative Action (Rural)	A								
	B								
Tier 1 - Affirmative Action (Minority)	A								
	B								
	C								
Tier 1 - Special majors	1								
Tier 2	A								
	B								
	C								
	D								
Tier 2 - Affirmative Action (Minority)	A								
	B								
	C								
Tier 2 - Special majors	1								
Tier 3	A								
	B								
	C								
	D								
Tier 3 - Affirmative Action (Minority)	A								
	B								
	C								
Tier 4 - Early Admissions	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 Deferred Acceptance (Parallel) Mechanism.

C.1.2 Measuring College Application Behaviors Using Choice Data

Based on features of the tier-specific applications in the Chinese centralized college admission system, we focus on three sets of strategies. These strategies are expected to capture some of the main application behaviors for a knowledgeable and skillful student. We have also covered these strategies in our interventions from the application guide "textbook", to school workshop, and to personalized advising. The first set of variables describe some general guidelines (or simple information/strategy):

- **[Strategy 1.1] 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 of variables.
- **[Strategy 1.2] 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 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 percentage 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 1.3] 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 admission 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 percentage 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.

The second set of variables describe the targeting strategies that students should use to apply to a combination of peer, reach/match and safety colleges (and majors). This strategy requires the most intensive knowledge and sophistication to make accurate predictions and decisions. This set of strategies are the key elements of our behavioral interventions as well as the data analysis in a student' 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 admission 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 et al., 2013). Besides, we focus on college-level application behaviors, but those choices of majors within each college are also worth exploring in future research.

- **[Strategy 2.1] Estimated gap (within 0.15 s.d.).** The behavioral rationale is that students should equate their CEE scores to admission scores in the previous years. 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 a CEE score of 500 should then look at colleges with admission scores around 550 in 2015. If she applied to colleges with admission 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 a CEE score of 600 in 2016 applied to colleges with admission 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.[44] 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.[45]
- **[Strategy 2.2] 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.

[44] 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.

[45] **??** shows that, 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.

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.[46]

- **[Strategy 2.3] Apply to colleges without admission data in the prior year.** The number of colleges that admit students in one province may change over time. Each year there are "new" colleges for students to apply to. The behavioral rationale is that students need to infer/predict the admission 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 admission data in the prior year. Sample mean is 2%.
- **[Strategy 2.4] 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* admission 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

[46] 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.

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 third set of strategies 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.

- **[Strategy 3.1] Minority affirmative action programs.** The behavioral rationale is that students may lack the information and knowledge to differentiate/understand different AA programs. National AA programs are of high quality (in selective colleges), but provincial AA programs are of lower-quality. We construct this variable by identifying whether a student applied to any AA programs. Sample mean is 22%, with a minimum of 0 and a maximum of 1.
- **[Strategy 3.2] 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 whether a student applied to any early admission programs. Sample mean is 15%, with a minimum of 0 and a maximum of 1.
- **[Strategy 3.3] 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 percentage of applied majors in teacher's education. Sample mean is 5.2%, with a minimum of 1 and a maximum of 40.

Student preferences and tastes are individual-specific and strictly unobservable. Particularly in constrained college applications, revealed preferences may not be precisely true. We construct three sets of proxy preferences using the application data. The first set includes college tuition and quota, which are the primary information provided to students by the Department of Education.

- **[Preference 1] College tuition and quota.** The behavioral rationale is that low-income students may prefer low-tuition colleges, and risk-averse students may prefer colleges with larger admission quota (Dynarski and Scott-Clayton, 2013; Hoxby and Avery, 2013; Loyalka et al., 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 admission 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 second set of preference variables are the college location choices:

- **[Preference 2.1] 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. 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 percentage of applied colleges locate 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 2.2] Out-of-province (advanced regions) colleges.** We construct this variable by calculating the percentage of applied colleges located in the most economically advanced regions of China, including Beijing, Shanghai, and 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 3] Majors.** We construct these variables by calculating the percentage 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.1.3 Correlations Between Applications and Admissions

Table 3 examines the correlations between applications behaviors and admission outcomes. Each application behavior group summarizes several variables within the group using a principal component analysis. In Table C.2, I report regression results using the itemized measures. The first two columns show the sample average in each measure between rural students and urban students, showing the poverty-gaps in college choice behaviors. Columns (4)-(7) add each group of strategy and preference variables stepwise. Consistent with the results in Table 3, among all the strategy and preference measures, targeting strategies explain the largest proportion of variations in college match.

Table C.2: College choices and the rural-urban gap in admission outcomes

		Sample mean		Outcome: Index of college match				
		Rural	Urban					
		(1)	(2)		(3)	(4)	(5)	(6)
# of colleges applied	Strategy 1	7.4	6.9		0.034***	0.019***	0.022***	0.022***
					(0.003)	(0.003)	(0.004)	(0.004)
# of colleges applied squared	Strategy 1				-0.001***	-0.001***	-0.001***	-0.001***
					(0.000)	(0.000)	(0.000)	(0.000)
% of majors applied	Strategy 1	67.3	73.1		0.001***	0.000	0.000*	0.000*
					(0.000)	(0.000)	(0.000)	(0.000)
% flexible major assignment	Strategy 1	63.4	76.4		0.000***	0.000***	0.000***	0.000***
					(0.000)	(0.000)	(0.000)	(0.000)
Estimated gap within 0.15 s.d. (=1)	Strategy 2	0.33	0.43			0.043***	0.046***	0.054***
						(0.008)	(0.008)	(0.008)
Did not apply for matched tier (=1)	Strategy 2	0.19	0.10			-0.526***	-0.515***	-0.518***
						(0.032)	(0.033)	(0.032)
Missing prior year data (=1)	Strategy 2	0.01	0.03			0.049***	0.049***	0.032**
						(0.012)	(0.012)	(0.012)
Descending (=1)	Strategy 2	0.27	0.41			0.097***	0.095***	0.082***
						(0.006)	(0.005)	(0.005)
Targeting (=1)	Strategy 2	0.32	0.45			0.035***	0.033***	0.031***
						(0.006)	(0.006)	(0.005)
Affirmative action (=1)	Strategy 3	0.29	0.19				-0.073***	-0.048***
							(0.013)	(0.012)
Early admissions (=1)	Strategy 3	0.16	0.17				0.046***	0.016*
							(0.010)	(0.009)
% teachers' colleges	Strategy 3	6.3	3.8				0.002***	0.002***
							(0.000)	(0.001)
College tuition (1000 RMB)	Preference 1	5812	6838					-0.024***
								(0.002)
College quota	Preference 1	860	518					-0.000*
								(0.000)
% out of province	Preference 2	29.3	50.6					0.001***
								(0.000)
% advanced regions	Preference 2	4.5	9.1					0.002***
								(0.000)
% economics majors	Preference 3	22.6	26.0					0.002***
								(0.000)
% agricultural majors	Preference 3	1.3	1.3					0.000
								(0.001)
% CS majors	Preference 3	2.9	3.5					0.004***
								(0.000)
% international majors	Preference 3	1.2	2.1					-0.002*
								(0.001)
% medical majors	Preference 3	12.8	9.7					0.001***
								(0.000)
Observations				28,806	28,806	28,806	28,806	28,806
R-squared				0.665	0.669	0.724	0.726	0.733

Notes: This table reports the OLS regression results of the correlations between college application behaviors and college match index, using data from those who submitted college applications in the untreated sample in 2016. Columns (1) and (2) report sample mean for rural and urban students. Regressions in columns (3)-(7) include a student's CEE score and other demographic covariates, as well as high school fixed effects. Standard errors in parentheses are clustered at high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

C.2. Intervention Descriptions (Guidebook & Workshop)

C.2.1 The guidebook

The “How to Apply to College" guidebook is to prepare all the relevant information and strategies that a student should have in the process of college choice and application. In 2016, we distributed the printed guidebook to treated students through high schools (on June 20). In 2017, we no longer distributed the printed version but used the electronic version for students in the “machine learning” advising group.

On the cover of the guidebook (Panels A and C in Figure C.1), we label that the guidebook is provided by a research team at Peking University (in 2016, as a joint team of Peking University and Ningxia University, the latter is the best college in Ningxia).

The outline of the guidebook is as follows (Panel D of Figure C.1):

1. Six steps in college applications
   - (a) Score equating
   - (b) Make use of past admission data
   - (c) Select a short list of colleges
   - (d) Identify the reach, peer and safety colleges and apply to a mixed set of them
   - (e) Major choices within each college
   - (f) Tier-specific plans (with a focus on the match tier)
2. Understanding college admission policies
   - (a) Background: Track, Tiers, Tier cutoff
   - (b) Deferred Acceptance (Parallel) mechanism
   - (c) College-then-major admissions
     - Major admission rules
     - Flexible assignment
     - Rejection and re-application
3. Supplemental materials
   - (a) Understanding the strategies of targeting reach, peer and safety college
   - (b) Useful information
     - Make use of your “advantages” (based on preference differentials)
     - Information and data collection

- Recommended online sources (Panel A of Figure C.2)
- National employment trends by majors (Panel B of Figure C.2)

(c) Application guidelines and tips

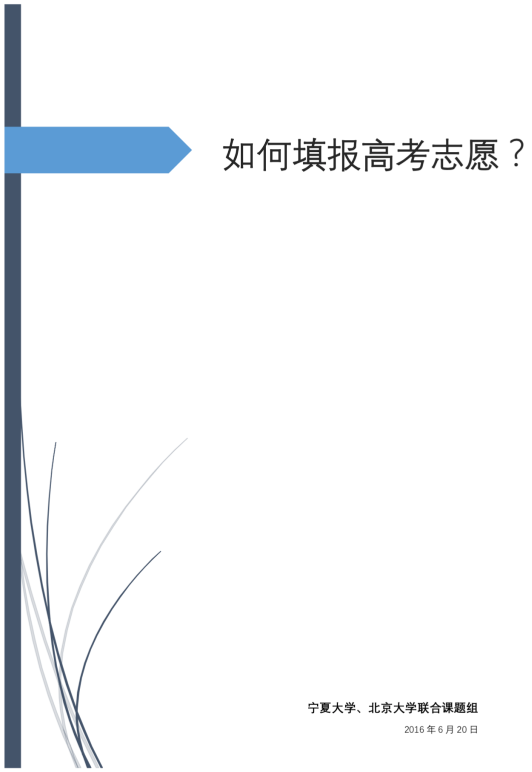
Recommended online sources. In preparing the guidebook, besides summarizing our own experience and knowledge, we have learned greatly from existing sources. Our research team carefully reviewed more than 200 Chinese websites and guidebooks that contained information about college entrance exam and college applications. We have also learned greatly from some excellent resources in the U.S., such as MDRC's "In Search of a Match: A Guide for Helping Students Make Informed College Choices" and the College Board's Big Future program.

In the guidebook, we provide a summary some of the most reliable and useful information to guide students to find the resources for further information. As shown in Panel A of Figure C.2, we list nine "college applications" websites, each of them covers some of the information that we think is relevant to college choices and applications. From left to right, these information items are:

- College introduction (1)
- Schools, majors within each college (2, 3)
- College admission guidelines (4)
- Admission scores (5)
  - The most reliable source is the printed book provided by the provincial Department of Education; we also purchased a few copies in 2016 and 2017 for the one-on-one advising
- Housing and dining (6)
- Recommended short list of colleges (7)
- Employment data (salary, locations; 8, 9)
- Degrees, major descriptions, coursework (10, 11, 12)
- Employment data (major-level salary, trends, locations; 13, 14, 15)
- Student evaluation (college, major; 16, 17)
- Major recommendation scores (18)

C.2.2 School workshop in 2016

We provided school workshop sin seven randomly chosen high schools. Workshops were organized by local districts and high schools. To minimize the quality variations in the workshops, we selected a group of very knowledgeable experts (editors of the guidebook) to give the workshop, using the same slides and scripts. Workshops were announced one month ahead of time in the name of a joint research team from Peking University (the top college in China) and Ningxia University (the top college in Ningxia). Each workshop lasted for three hours and was moderated by a high-level school administrator. Figure C.3 and Figure C.4 show the sample pictures.



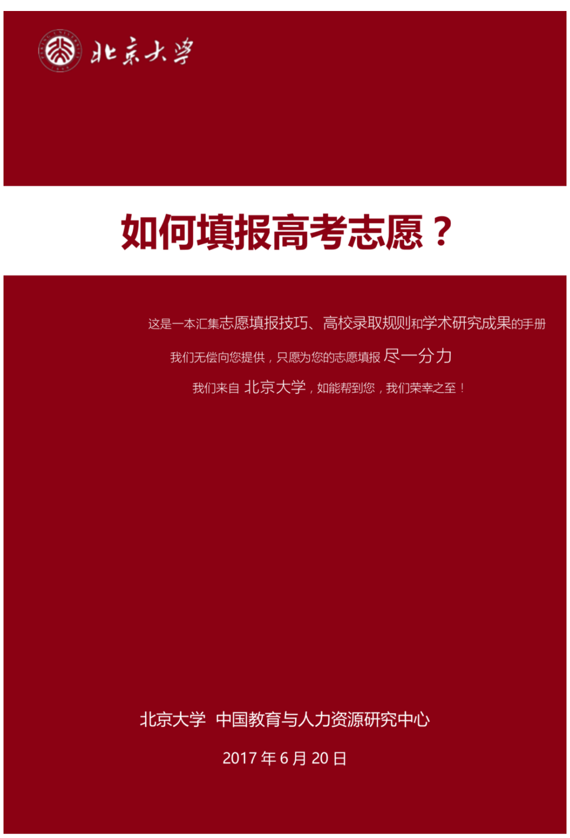
如何填报高考志愿？

宁夏大学、北京大学联合课题组

2016 年 6 月 20 日

(a) Cover (2016 edition)

(b) Received packages from the press, 2016, Beijing



北京大学

如何填报高考志愿？

这是一本汇集志愿填报技巧、高校录取规则和学术研究成果的手册

我们无偿向您提供，只愿为您的志愿填报 尽一分力

我们来自 北京大学，如能帮到您，我们荣幸之至！

北京大学 中国教育与人力资源研究中心

2017 年 6 月 20 日

(c) Cover (2017 edition)

目录

一 高考志愿填报基本步骤 1

第一步：做好高考分数“等值化”，找准自己的位次！........ 2
第二步：关注往年录取趋势，看清录取的“潮流”！........ 3
第三步：综合实际兴趣，明确意向高校的范围！........ 4
第四步：意向高校的归类和排序，采用“冲保”结合策略！........ 5
第五步：专业志愿的取舍和抉择，不能盲目“冲”专业！........ 6
第六步：明确志愿填报“批次”，避免不合意地“错”录！........ 7

二 高考志愿填报规则解惑 8

第一篇：基础知识 9
考生类别 9
批次划分 9
批次线 10
第二篇：平行志愿 11
平行志愿投档与高校录取规则 11
录取细节 11
误区 12
平行志愿面面观 12
第三篇：专业录取 13
专业志愿 13
专业录取规则 13
调剂 13
退档与“征集志愿” 14

三 补充说明和实用材料 15

第一点：冲保策略的内在逻辑 16
第二点：实用材料 18
志愿填报中“优势”的把握 18
资料的收集和整理 18
院校和专业信息获取途径整理 19
往年分专业高校毕业生就业状况及趋势概览 20
第三点：写在最后的“实用策略” 27

I

(d) Outline (2017 edition)

Figure C.1: The guidebook “How to Apply to College?”

Notes: This figure shows sample pictures of the guidebook in 2016 and 2017.

表格 1 志愿信息参考网站一览

网站名称、网址	大学情况	院系设置	所设专业	招生章程	分数线	生活条件	估分推荐	学校就业起薪	学校就业去向	专业学位划分	专业培养目标	专业核心课程	专业就业起薪	专业就业趋势	就业方向	院校满意度	专业满意度	专业推荐度
新浪教育·高考院校库 http://kaoshi.edu.sina.com.cn/	√		√	√	√	√	√	√			√	√						
中国教育·在线高考志愿填报系统 http://gkcx.eol.cn/	√		√			√	√				√	√						
搜狐教育·搜狐大学信息库 http://daxue.learning.sohu.com/	√		√		√		√		√									
看准网·大学专业 http://www.kanzhun.com/dxjy/											√	√	√	√	√			
高考网·专业信息 http://college.gaokao.com/spelist/										√	√	√			√			
高三网·大学专业解读 http://www.gaosan.com/zhuanyejiedu/											√			√	√			
学信网·阳光高考 http://gaokao.chsi.com.cn/	√	√	√	√		√										√	√	√
第一高考网·找专业 http://www.diyigaokao.com/major/bklist.aspx											√		√		√			
中国教育在线 http://www.eol.cn/html/g/benkezy.shtml										√	√	√						

3

(a) Sumamry of reliable online resources

表格2 分专业本科毕业生规模结构与初次就业率

学科专业		2011年规模与结构		2011年初次就业率（%）	11-14年初次就业率变化趋势
学科门类	专业类	人数	百分比		
哲学	哲学类	2172	0	85	
经济学	经济学类	173853	6	88	
法学	法学类	74326	3	79	
	马克思主义理论类	193	0	90	
	社会学类	11424	0	85	
	政治学类	20058	1	84	
	公安学类	11288	0	77	
教育学	教育学类	36637	1	83	
	体育学类	55046	2	78	
	职业技术教育类	8621	0	91	

(b) Trends in employment rate by majors

Figure C.2: Sample contents in the guidebook “How to apply for college?”

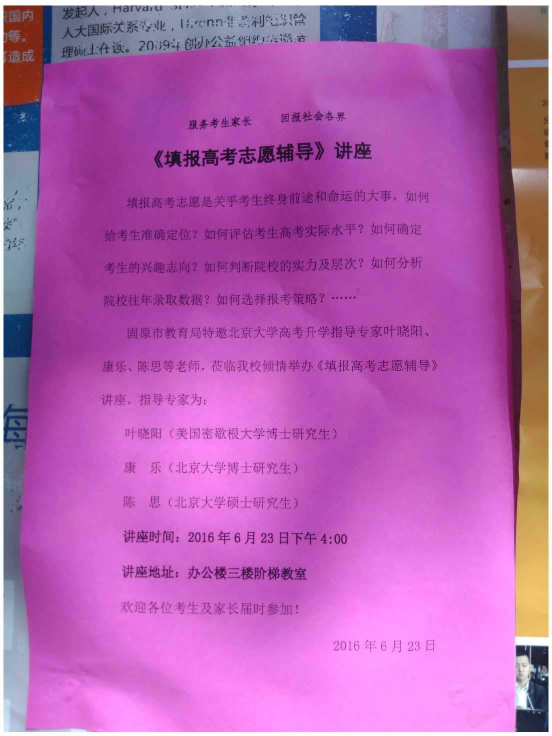
Notes: This figure shows sample contents in the guidebook. Panel A lists nine websites with a cross-tab of available information on each website that we selected from about 200 Chinese websites. Panel B shows that the employment trend graph by major that was created using data on every college graduate from 2011 to 2014.

(a) Guyuan No.1 High School (Speaker: Xiaoyang)

(b) Helan No.1 High School

Figure C.3: High school workshops in 2016

Notes: This figure shows sample pictures of the school workshops in 2016.



(a) School poster

(b) Q&A after workshop

Figure C.4: High school workshops in 2016 (Guyuan No.2 High School)

Notes: Figure A shows the school poster. The workshop was announced as organized by Guyuan City Department of Education. Figure B shows the brief conversations with students and parents after the three-hour workshop. The sentence on the back of the project tee "**Only the educated are free**" is from a Greek Stoic philosopher Epictetus (AD 55-135). While each workshop had one speaker, we had a team of 3-4 members there for brief follow-up Q&A after each workshop.