Improving College Choice for the Poorest Students Using Behavioral Policy Interventions

by

Xiaoyang Ye

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Higher Education)
in The University of Michigan
2019

Doctoral Committee:

Professor Susan M. Dynarski, Co-Chair
Professor Brian A. Jacob, Co-Chair
Professor Stephen L. DesJardins
Professor Brian P. McCall
Professor Jeffery A. Smith, University of Wisconsin-Madison

Xiaoyang Ye

yxy@umich.edu

ORCID iD: 0000-0003-2872-824X

DEDICATION

To

Samuel Pollard (1864-1915)

Y. C. James Yen (1890-1990)

Ronald Yu Soong Cheng (1899-1992)

Li Zhi (1978-)

For

Their faith, struggle, sacrifice, enlightenment

In expanding educational opportunities for the poorest people

奮斗與犧牲務把文盲除盡

ACKNOWLEDGMENTS

"Only the educated are free."

"You don't have to say a word."

We human beings often think that living for a long time is a long life. But we don't know how many wrinkles can be folded in a second. Much of my life has been a white pony flashing by while sprinkling scores of glowing fireflies into the universe. It has been a blessing to have all the people behind me or above the sky. One of the motives for my Ph.D. studies is the authorization to write the Acknowledgments section. This is for them.

First and foremost, I gratefully acknowledge the six years under the essential guidance, mentoring, and support from my dissertation committee: Stephen DesJardins, Susan Dynarski, Brian Jacob, Brian McCall, and Jeffrey Smith. Six years ago, I decided to choose Michigan over other options because all of them are here (Jeff, you were here). I aspired - or maybe in an appropriate word, dreamed - to be a future education policy researcher like them with a dedication to scientifically rigorous research for important real-world problems. I am extremely grateful to the opportunity that I have managed to have them on my committee. I hope what I have learned from them, what I have done in the research with them or independently, and what I plan to do in the coming years with them or independently would make me "an officially qualified student" of theirs. I also hope all those things would theoretically and empirically demonstrate that I made an optimal life choice to come to Ann Arbor, as well as that they made a "non-bad" choice to have me as a student.

Susan Dynarski is famous! It is partly because of her Does Aid Matter paper, and I got to know that paper ten years ago. On December 24, 2009, two days after returning to Beijing from volunteer-teaching in a village school in Northwestern China, I worked overnight to finish the review report of the Aid paper for my senior thesis course. In my early twenties, when I thought Soccer Weekly much more attractive than AER, I did not know what would happen to me in the next ten years. But, as it happened, fortune stood me.

Brian Jacob is not less famous. Nearly one year later, I read Brian's 884-cited Accountability paper in my first-year master's Economics of Education class. This paper was the first to motivate my long-term interests in studying incentives in education policy. Without counterfactuals, I cannot assert that I would not have stepped down the Ph.D. journey (vs. a business career as I can easily see the average in my close friends) in the absence of encountering these two great papers early in my career. But when it came to a choice between Michigan and other places, I saw the answer is within the papers.

Sue and Brian have been the best dissertation co-chairs that I could hope for. As one of the youngest members of the army that they have trained, I have gained every expertise that is needed for conducting the modern cutting-edge research as well as for teaching the new generation of education policy professionals as they do. Their guidance has helped me get through many highs and lows: they care about my study, my research, my family, and my life. In the summer of 2014, I wrote to accept the continuing research assistant position: yes, I'd like to continue in the following year, and many years. They replied: not TOO many years, we want you to graduate and feed your family! Not too many years later, I made it!

I have been lucky enough to join a second wonderful research group at Michigan, led by Stephen DesJardins and Brian McCall. Steve interviewed me during the doctoral program application and brought me to Michigan. We pleasantly extended the 30-minute interview into a 2-hour conversation. For the past six years, Steve has spent countless hours helping me in every possible way. As Ozan wrote in his dissertation acknowledgments, Steve fights for us behind our back: he guided my coursework training, provided research collaboration opportunities, supervised my teaching of the advanced quantitative course in which I experienced the largest value-added in my teaching interests and skills, taught me to be professional in the field of higher education, and babysit my year-long job search. I appreciate that Steve always pushes me to think deeper about the theories and contexts in higher education to bridge the border between economics and education. Brian has offered much of his generous guidance and support together with Steve. I first met Brian in Beijing when we were at the 2012 International Conference of Economics of Education. My research at that time focused on K-12 education, but I was quite impressed by Brian's (& Steve's) research. I enjoyed the collaboration with Brian and Steve in the past years on high school coursetaking and college choice. I also thank Brian for taking care of my first year and particularly for recommending me a course sequence of formal economics training.

I started to interact with Jeffery Smith in his doctoral-level Applied Econometrics class and the CIERS seminars. Before Jeff left Ann Arbor, he attended all my CIERS presentations. On December 29, 2015, Jeff sent me a long list of comments on my descriptive paper of college undermatch in China that I presented in September - much longer than all the referee reports that I have received for that paper in field top journals. Though all my committee members are world-class economists, Jeff has played a statistically and substantially significant role in keeping me methodologically correct in my papers. I also like his personal

stories in the economics profession that Jeff shared in the Causal Inference in Education Policy class. Jeff is exceptionally generous with his time. He made his time available to meet whenever he came back to Ann Arbor. I greatly appreciate his care and encouragement for my job search. It was great to give an official job talk when a committee member was in the audience to show his support! I hope I could do it once again in the near future.

Many professors at Michigan have provided valuable conversations and support that resulted in this dissertation, including John Bound, Phillip Bowman, Isaac McFarlin, Natasha Pilkauskas, Kevin Stange, and Christina Weiland. I also thank Peter Bahr, Mike Bastedo, Lisa Lattuca, Jan Lawrence, Vilma Mesa, Julie Posselt, Awilda Rodriguez, Ed Silver in the School of Education for their help. Jessica Mason, Linda Rayle, and Melinda Richardson were always ready with answers to the messy administrative questions. As an "empirically-minded Stata runner," I very much enjoyed the qualitative research class by Pamela Moss. Special thanks to John DiNardo for having shaped my mind of empirical research and, more importantly, for showing me how to laugh at and enjoy life, no matter how tough it is.

I have benefited tremendously from many other professors in our field. I thank Dom Brewer for the Ph.D. admission offer and his continuous support over the entire course of my Ph.D. journey. I thank Siwei Cheng, Sean Corcoran, Will Doyle, Avi Feller, Lindsay Page, Laura Perna, Yu Xie, and Liang Zhang, all of who provided very helpful conversations that influenced the dissertation chapters and my academic career in general. I am particularly grateful to Xiaogang Wu for initiating me into rigorous empirical research. Thanks also go to Richard Murnane, Karthik Muralidharan, and Noreen Weeb for their exceptional mentoring during the NAEd/Spencer Dissertation Fellowship. Special thanks to Michal Kurlaender for her great advice and help in the past few years.

I thank Prashant Loyalka, my co-author for many papers, for his enthusiasm for Chinese education policy problems, for being a role model of a great researcher, and for his advice and close friendship throughout my doctoral studies. Our first paper on college undermatch directly motivated the Bright Future of China Project that resulted in my dissertation. Thanks also go to Michelle and the Loyalka kids for warm hosting during my visits to Stanford. Scott Rozelle and the REAP team also provided a welcoming research environment and insightful conversations.

Perhaps I could never find a better intellectual home than the Education Policy Initiative. I spent nearly 20,000 hours on the fifth floor of Weill Hall and saw many people come and go in the past six years. Most importantly, we worked together to improve education for all. The staff team is phenomenal: Julie Monterio de Castro, Tedi Engler, Mahima Mahadevan, Nicole Wagner, DeVante Rollins, Jasmina Camo-Biogradlija, and Ramona Avram were very reliable and endless sources of administrative assistance. EPI has gathered many brilliant young education policy scholars, and I have been lucky enough to work with multiple cohorts and enjoy the unique friendship. Working with them in these years, I have greatly benefited from learning from all of them. An incomplete list includes Meredith Billings, Sarah Cannon, Fernando Furquim, Thomas Goldring, Max Gross, Steve Hemelt, Monica Hernandez, Jonathan Hershaff, Max Kapustin, Dan Kreisman, CJ Libassi, Andrew Litten, Elizabeth Mann, Shawn Martin, Kathy Michelmore, Meghan Oster, Stephanie Owen, Silvia Robles, Rachel Rosen, Anna Shapiro, Andrew Simon, Brittany Vasquez, Mark Wiederspan.

I owe special thanks to Josh Hyman, Daniel Hubbard, and Julian Hsu. Josh taught me to get started with all the Michigan data in my first year and has provided me many guidance and support ever since that. He also largely improved the quality of my job market paper. I spent four years with Daniel, my best office mate so far, in the same office before he graduated. We shared numerous memorable moments. Julian has offered many valuable Ph.D. and career advice. I enjoy our every meeting at AEFP or APPAM. Outside EPI, Ozan Jaquette and Richard Welsh are my truly respected elder brothers and I always look forward to the next drink that we will have somewhere across the country. All these gentlemen are conducting first-tier research on education policy. But I am lucky enough to benefit from them no less than those kids who would be impacted by their research.

CSHPE is the other intellectual home for me at Michigan. As the only student in the 2013 cohort, I was fortunate to be welcomed as an extra member to join the 2012 cohort from the summer before I arrived in Ann Arbor (Kyle and Jon emailed me to offer a ride two months before my flight). I thank Michael Brown, Matthew DeMonbrun, Feven Girmay, Esmeralda Hernandez, Grant Jackson, Jon McNaugthan, Jennifer Pollard, Kimberly Reyes, and Kyle Southern for their friendship, support, and guidance throughout the journey. I also thank CSHPE friends in other cohorts, including Yiran Chen, KC Deane, Kristen Glasener, Jeff Grim, Aurora Kamimura, Heeyun Kim, and Sooji Kim. Wenting Lan and Peilin Qiu provided great help with my defense meeting and other graduation issues. Zheng Cui spent several joyful years with me in the program.

Have spent 29 years in school as a student, I am indebted to many people who provided education, advice, support, and friendship. Before looking back in the past, I wish to thank Adam Kapor, Christopher Neilson, and Seth Zimmerman for hosting my next stop at Princeton, and Susanna Loeb for the position at Brown. I look forward to the fascinating research projects with them in the coming years.

All my research would not be possible if I were not fortunate enough to have Yanqing Ding and Rong Wang, two of the best economists of education in China, as my advisors in my master's program. They laid all the foundations for being a qualified education policy researcher during my transition from a poet working with words to an economist working with numbers. They also inspired me and set me on the path to work hard for the Ph.D. degree 10,000 miles away from China. I have been working closely with Yanqing for a decade, particularly the huge-scale *Bright Future of China Project*. Rong opens all the possible doors for me to explore the complex education finance issues in China. I am still far from their expectations, but I hope I could catch up in the next ten or twenty years. Wei Ha has been a very close co-author, mentor, and friend. I also thank professors in the Graduate School of Education at Peking University who not only supported my master's studies when I was there, but also continue to stay in good touch in the past years, including Wei Bao, Xiaoyu Chen, Xiaohao Ding, Mingxing Liu, Liping Ma, Yingquan Song, Zhilei Tian, Jianguo Wei, Fengqiao Yan, Po Yang, and Changjun Yue; as well as those earlier GSE graduates, including Shuangye Chen, Dan Mao, Haiping Xue, and Suhong Yang. Bingyu Sun has taken care of all the administrative stuff. My undergraduate advisors in the School of Economics, Shuanglin Lin, Yi Liu, Dashu Wang, and Cheng Yuan introduced me to public economics and have provided continuous encouragements in the past 12 years. Without the countless help from my high school teachers Mingyong Chen, Wuhong Han, and Junchang Li from nearly twenty years ago, it was impossible for me to attend the top colleges in China and the U.S.

Many people have contributed to the *Bright Future of China Project* that resulted in this dissertation. Thanks to Wei Li and Yujie Ma for their hard work in the field; Keqiang Li, Tzuyi Yu, Jingxin Wang for the technical assistance; and Dongyang Chen, Wang Fu, Kailing Guo, Yifan Wang, Yu Xiong, Haiyan Yang, and Yongxin Ye for their superb assistance. Members of the Ding Lab have been the key personnel in the project, including Si Chen, Lizhen Du, Honghui Li, Yafei Lian, Yingjie Lin, Xiao Liu, Qiongqiong Wang, Xuanheng Wu. Yuchuan Wu traded his academic talent for management expertise, and finally, I got the opportunity to work with him. Jing Liu is the best project manager that I have found. Fengming Lu, Le Kang, and Hongbin Wu all become my "excited" friends and collaborators. It was the greatest loss in my life that Jin Yang and Yinduo Wu decided to build the Chinese socialism in business and political affairs. I am always waiting for them to be back: we have 100 papers to write before alleviating the universal poverty by education.

A few more thanks: I thank Shaoda Wang, my most gifted co-author, for our long-term collaboration since 2013; and together with Ao Wang for the third chapter. I thank the Chinese scholars in our field for their support and the great gatherings at conferences. It feels good to be part of this group. Thanks to Li Feng, Wei Li, Yifeng Luo, Yongmei Ni, Xu Qin, Florence Ran, Yang Song, Yi Wei, Qiao Wen, Di Xu, Rui Yang, Yuan Zhang, Sen Zhou, and Qiong Zhu. Jing Liu and I graduated from the same undergraduate and master's programs and are now working in the same field. Thanks to Jing for sharing moments in graduate school, research, and the hotel room in every conference we attended together.

Years ago, I had quite higher reputations as a Chinese poet than as an economist of education. I particularly thank Bo Han, Xudong Hu, Tao Jiang, Di Zang, Shu Zhou for their guidance. As John Keating said in Dead Poets Society, we read and write poetry because we are members of the human race. I am very grateful to my poet friends, Zhichen Cong, Xue Fan, Buyan He, Xingzhou Hu, Yong Jin, Yin Liu, Yiwen Liu, Yuanhang Liu, Min Peng, Jianqiang Wang, Long Wang, Qinlin Wang, Xiangting Wu, Chao Xia, Yue Xu, Yang Yu, Li Zhang, Qiong Zhang, and Yijing Zheng at Peking University Wusi Poetry Society, and Sen Cao, Di Cheng, Bubai Gu, Xiaojian Li, Zhan Luo, Shou Mu, Dale Wang, Shui Xiao, Xiao Xu, and Dan Ye. Xiangjun Feng is the best poet in the history of Peking University.

My time in Ann Arbor was much happier due to the excellent company provided by Ziwei Cao, Tangren Feng, Xing Guo, Ting Lan, Meizi Li, Shuqiao Sun, Yuchen Yang, Hang Yu, the Chinese Soccer Team at the University of Michigan, the PKU Weiming Soccer Team, and Michigan Table Tennis Club. Xinwei Ma is the best econometrician of our time.

Beyond the college town life, I thank the long-term friendship provided by Peng Bi, Xuelian Cao, Pi Wen Chen, Binglong Li, Zhao Liu, Yajing Qian, Zheng Wang, Zhifei Xu, Rui Yao, Chenhui Zhang, and Kaiwei Zhang from graduate school; Jie Feng, Yinshen He, Junxiang Hu, Xiao Hu, Yu Jin, Qingwen Lan, Miao Lei, Jiaxuan Li, Yifeng Su, Zhe Sun, Yu Xu, Ke Zhang, and Yifan Zhang from college; and Changxiao Huang, Bo Hou, Chen Li, Han Pu, Ye Xiong, Dong Zeng, Chenming Zhang, Han Zhang, Linli Zhang, Ruyue Zhang, Xiao Zhang, Yi Zhang, Mi Zhou, and Nan Zou from high school. I spent every night with the music by Li Zhi, Fantong, Liu Dongming, Su Yang, and Lin Shengxiang. Yajing is the best musician born in Chongqing. Though I can write another dissertation for this, special thanks to two ten-year-old groups Fanfou and SYDT for every day and night, near and far.

Thank you to my family for unconditional love. My parents have always supported me in whatever I do and whatever they can. I hope they are proud of me at this moment. My cousin Yao Che merits one sentence here as he has helped with everything when I am not at home. Being the one-child generation in China, Mo Guo and Yang Yu are the best brothers that I could hope for. Finally, my son, you motivate me to strive for education policies that work best for all the very individual kids. It all began the day I met your mother.

I sometimes ask myself the same question repeatedly. If I suddenly died on the 58-year-old weekend like Alan Krueger, what should I do for the next 27 years? Admittedly, the first task is to find what I want to do. And then, work hard and bravely for it. Life is too short, like sailing against the current. It is not a big deal if we have all become graves. But, I still cannot stop thinking: what should I do to revitalize the future of mankind?

I too will wag my tail in the mud.

(June 4/July 30, 2019. Ann Arbor.)

PREFACE

For the first two chapters, I thank Joshua Hyman for carefully reading the draft and providing feedback. Besides my dissertation committee, I benefited greatly 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, 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, Rong Wang, Shaoda Wang, Noreen Webb, Christina Weiland, Po Yang, Cheng Yuan, Liang Zhang, and participants at Advances in Field Experiments, AEFP, AERA, APPAM, ASHE, NAEd Retreat, Michigan (CIERS, Education, Ford), Peking University (Economics, Education), RAND, Tsinghua University, University of Florida, University of Washington, University of Wisconsin-Madison, and an anonymous reviewer at the ASHE conference.

For the third chapter, Ao Wang, Shaoda Wang and I thank Ned Augenblick, Aaron Bodoh-Creed, Celeste Carruthers, Ernesto Dal Bo, Stefano DellaVigna, Susan Dynarski, Marco Gonzalez-Navarro, Brian Jacob, Supreet Kaur, Lixing Li, Ethan Ligon, Fengming Lu, Xinwei Ma, Jeremy Magruder, Aprajit Mahajan, Sakib Mahmood, Alex Rees-Jones, Elisabeth Sadoulet, Reed Walker, Christina Weiland, Yang Xie, David Yang, Noam Yuchtman, and numerous seminar and conference participants for helpful discussions. Xin Li and Qiong Zhu provided excellent assistance with the fieldwork.

I particularly wish to thank Yanqing Ding (Peking University, the other co-PI of the Bright Future of China Project), Wei Li (Ningxia University), Bingqing Dai, and Wei Geng (Ningxia Department of Education) for their extraordinary effort in initiating the project, Xiaoyang Hu, Jing Liu, 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 would like to 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 these chapters. The Bright Future of China 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, the National Natural Science Foundation of China (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.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGMENTS iii

PREFACE ix

LIST OF TABLES xiv

LIST OF FIGURES xvii

ABSTRACT xix

CHAPTER

I Improving College Choice for the Poorest Students: Results from Randomized Experiments in Centralized Admissions 1

1.1 Introduction 1
1.2 Background 7
1.2.1 Theoretical Framework: Information and Knowledge in College Choice 7
1.2.2 Centralized College Admissions 11
1.2.3 Context: College Choice in China 14
1.2.4 The Bright Future of China Project 18
1.2.5 Data Sources 19
1.3 The Policy Challenge 20
1.3.1 College Undermatch in Centralized Admissions 20
1.3.2 Poverty Gap in College Undermatch 24
1.3.3 The Potential for Behavioral Interventions at Scale 26
1.4 Experimental Design 29
1.4.1 Interventions: Promoting the Learning of Knowledge in College Choice 29
1.4.2 School Level Randomization 33

1.4.3 Econometrics . . . 35
1.5 Results . . . 36
1.5.1 Effects on Admissions Outcomes . . . 36
1.5.2 Effects on College Application Behaviors . . . 40
1.6 Summary . . . 42
1.7 Main Figures . . . 45
1.8 Main Tables . . . 49
1.9 Additional Figures . . . 56
1.10 Additional Tables . . . 61
1.11 Additional Descriptions . . . 70
1.11.1 Correlate College Application Behaviors with College Admissions . . . 70
1.11.2 Intervention Descriptions (Guidebook & Workshop) . . . 82
1.11.3 Intervention Descriptions (One-on-One Advising) . . . 89
1.11.4 Estimating the Effects of the One-on-One Advising in the 2016 Program 93

II Human Instruction vs. Machine Learning: Experimental Evidence on Improving College Access and Match at Scale 98
2.1 Introduction . . . 98
2.2 Background . . . 107
2.2.1 Why Do Some Students Make Suboptimal College Choices? . . . 107
2.2.2 Why Do Centralized Admissions Need (More) Strategic Applications? 109
2.2.3 Why Did We Create the Bright Future of China Project? . . . 112
2.3 Conceptual Framework . . . 114
2.4 Interventions . . . 118
2.4.1 Intervention for Students: Machine Learning . . . 118
2.4.2 Intervention for Teachers: Pay for Performance . . . 122
2.5 Experimental Design . . . 125
2.5.1 Data Sources and Motivating Evidence . . . 125
2.5.2 Experimental Design: Individual Level Randomization . . . 129
2.5.3 Econometrics . . . 134
2.6 Results . . . 136
2.6.1 Effects of Machine Learning Assisted Advising . . . 136
2.6.2 Thinking Fast? Slow! . . . 143
2.6.3 Effects of Teacher Incentives . . . 146
2.7 Discussion and Policy Implications . . . 150
2.8 Conclusion . . . 153

2.9 Main Figures 155
2.10 Main Tables 161
2.11 Additional Figures 170
2.12 Additional Tables 175
2.13 Additional Descriptions 189
2.13.1 Measuring College Application Behaviors Using Actual Choice Data . 189
2.13.2 Intervention Descriptions (One-on-One Advising) 197
2.13.3 Intervention Descriptions (Teacher Pay-for-Performance) 200

III Religion and Motivated Cognition: When Ramadan Meets the College Entrance Exam 202
3.1 Introduction 202
3.2 Background 208
3.2.1 The College Entrance Exam in China 209
3.2.2 The Muslim Ramadan Fasting 210
3.2.3 Ramadan and Exams 212
3.2.4 The Costs of Taking the CEE During Ramadan 214
3.3 Experimental Design and Implementation 216
3.4 Testable Hypotheses 221
3.5 Results 224
3.5.1 Students' Priors on the Cost of Ramadan 225
3.5.2 Existence of Motivated Cognition and Effects of Exemption 226
3.5.3 Mechanisms 227
3.5.4 Fasting Decisions 232
3.6 Conclusion 234
3.7 Figures 237
3.8 Tables 242
3.9 Setup of the Theoretical Model 253
3.10 Reading Materials 260
3.10.1 Reading on Exemption - The Treatment Condition 260
3.10.2 Reading on Art - The Control Condition 261

BIBLIOGRAPHY 264

LIST OF TABLES

TABLE

1.1 Experimental design: Interventions in 2016 49
1.2 Extent of academic undermatch: College access vs. college choice 50
1.3 Poverty gap in college access and match 51
1.4 College choices and the poverty gap in admissions outcomes 52
1.5 Balance checks: 2016 RCT 53
1.6 ITT effects on college access and match outcomes: Guidebook and workshop in 2016 54
1.7 ITT effects on college choice behaviors: Principal-component factors 55
1.8 Measures of undermatch: Varying thresholds 61
1.9 Rural-urban gap in college access and match: Itemized outcomes of the index measure 62
1.10 Rural-urban gap in college access and match: Additional outcomes 63
1.11 Sample description: 2016 RCT 64
1.12 Balance checks: Prediction of treatment status using student-level covariates in 2016 RCT 65
1.13 ITT effects on additional college access and match outcomes: Guidebook and workshop in 2016 66
1.14 ITT effects on college access and match outcomes for high achieving students: Guidebook and workshop in 2016 67
1.15 ITT effects on college choice behaviors: Itemized results 68
1.16 ITT effects on college choice behaviors for high achieving students: Principal-component factors 69
1.17 College application form in Ningxia in 2016 (Simplified) 72
1.18 College choices and the rural-urban gap in admissions outcomes 81
1.19 Take-up of the individualized advising program 96
1.20 Effects of the individualized advising program 97

2.1 Experimental design: Interventions in 2017 161

2.2 College choices and the poverty gap in college match 162
2.3 First stage: Take-up of individualized advising programs 163
2.4 ITT and TOT effects on college access and match of the individualized advising programs 164
2.5 ITT and TOT effects of individualized advising programs: College application behaviors 165
2.6 Heterogeneity in the ITT effects on application time use 166
2.7 ITT effects of teacher incentives: College access and match outcomes 167
2.8 Measuring teachers' belief, knowledge and action in college application advising 168
2.9 Heterogeneity in the ITT effects of teacher incentives: By teacher characteristics 169
2.10 The poverty gap in college match 175
2.11 Balance checks for student interventions 176
2.12 Balance checks for student interventions: Prediction of treatment status 177
2.13 Balance checks for teacher interventions: Teacher-level covariates 178
2.14 Balance checks for teacher interventions: Student-level covariates 179
2.15 Balance checks for teacher interventions: Prediction of treatment status 180
2.16 ITT and TOT effects of individualized advising programs: Itemized outcomes in the index measure 181
2.17 ITT and TOT effects of individualized advising programs: Itemized college application behaviors 182
2.18 Heterogeneous ITT effects of individualized advising programs 183
2.19 Time use in college applications 184
2.20 Correlations between application time use and outcomes in college choices and admissions 185
2.21 ITT effects of teacher incentives: Itemized outcomes in the match index 186
2.22 ITT effects of teacher incentives: College application behaviors 187
2.23 Heterogeneity in the ITT effects of teacher incentives: By student characteristics 188

3.1 Impacts of Ramadan on CEE score 242
3.2 2*2 experimental design 243
3.3 Balance tests 244
3.4 The effect of exemption under unawareness 245
3.5 The effect of exemption on graph reading (Hui-Han CEE gap) 246
3.6 Extensive margin adjustment 247
3.7 Heterogeneity based on fasting history 248
3.8 Heterogeneity based on valuation for college 249

3.9 Effect of exemption on graph reading (GDP per capita) 250
3.10 Active information distortion 251
3.11 Fasting attitudes 252

LIST OF FIGURES

FIGURE

1.1 Rural-urban gap in academic undermatch 45
1.2 Correlates of college choices and admissions outcomes 46
1.3 Heterogeneity in the ITT effects: All students 47
1.4 Distributions of college peer median score between students who received 1-on-1 advising and students in the control group in 2016 48
1.5 Project timeline 56
1.6 Location of Ningxia 57
1.7 Testing the sensitivity of different college quality measures 58
1.8 Distribution of the distance between college median score and a student's own score 59
1.9 Heterogeneity in the ITT effects: High achieving students 60
1.10 The guidebook "How to apply for college?" 85
1.11 Sample contents in the guidebook "How to apply for college?" 86
1.12 High school workshops in 2016 87
1.13 High school workshops in 2016 (Guyuan No.2 High School) 88
1.14 Example of the online individualized advising in 2016 92

2.1 The poverty gap in academic match 155
2.2 How does *Stata* reduce human labor? 156
2.3 How does machine learning work in 2017? 157
2.4 Distribution of college application submission time 158
2.5 Teacher survey responses 159
2.6 Heterogeneity in the ITT effects by student CEE score distribution 160
2.7 Project timeline 170
2.8 Location of Ningxia 171
2.9 Relative importance of features in the random forest predictions 172
2.10 Example of the online individualized advising in 2017 173
2.11 Distribution of teachers' value-added of college-going advising 174

2.12 Teacher pay-for-performance information letter in 2017 200
2.13 Teacher pay-for-performance contract letter in 2017 201

3.1 Hui-Han score gap in the College Entrance Examination (2011-2016) 237
3.2 Event study estimates of the Hui-Han CEE gap 238
3.3 Sino-Japanese income gap (2011-2016) 239
3.4 Distribution of guesses of the enlarged Hui-Han CEE gap in 2016 240
3.5 Mean agreed response items for the list experiment across treatments 241

ABSTRACT

Inequality in college access and match is a persistent global problem. Only recently, misinformation and sub-optimal decisions have been recognized as major behavioral barriers for low-income students during the complicated high school to college transition. However, little is known about what works in centralized college admissions systems or developing countries. To fill this gap, from 2016 to 2018, I designed and conducted the *Bright Future of China Project*, a set of large-scale randomized experiments, in one of the poorest provinces of China (Ningxia). This three-chapter dissertation contributes to the literature and worldwide policy efforts by providing new evidence on using behavioral policy interventions to improve college access, choice, and match at scale for low-income students.

The first chapter studies college choice behaviors and admissions outcomes in a typical centralized system in China. I document that the student-college academic undermatch is prevalent in centralized college admissions, even though the application process is simplified, and college information is centrally provided. I find descriptive evidence that the undermatched college choices are likely due to the lack of accurate predictions of admissions probabilities. I present results from school-level randomized experiments in 2016, which targeted students' behavioral barriers by providing college application assistance in the forms of a guidebook, a workshop, and personalized advising. The intensive application assistance interventions, during a short period with low costs, largely improve college access and match by shaping students' college choice behaviors.

The second chapter studies the scale-up problem of behavioral interventions in college choice for low-income students. I report results from a set of student-level and teacher-level

randomized experiments in 2017. Treated students were provided with a personalized advising program that targeted their behavioral barriers in the process of college choice and application. To scale up the labor-incentive advising, I designed and examined two policy solutions: (1) machine learning algorithms to simplify predictions and improve decision-making, and (2) a pay-for-performance policy to incentivize teachers to act as temporary counselors. Machine learning substantially increased advising effectiveness. However, without complementary school organizational policies, teacher pay-for-performance policy did not incentivize teachers to provide sufficient application assistance to students.

The third chapter, co-authored with Ao Wang and Shaoda Wang, examines how motivated cognition prevents students from learning crucial objective information, using a field experiment in 2018. We focus on a unique empirical setting: Chinese Muslim students were about to take the high-stakes College Entrance Examination during the month of Ramadan, due to the occasional overlap of these two events. We invited well-respected Chinese Muslim leaders to grant explicit exemptions to delay the Ramadan fast until after the exam. We then randomly provided this treatment to some Muslim students, creating experimental variation in the stringency of religious practices. Using a survey design, we measured students' perceived cost in exam performance of taking the exam during Ramadan. Comparing with the estimates that we obtained from a novel administrative dataset and an event study design, we find that the control group students who did not receive the exemptions exhibited strong patterns of motivated cognition bias. They underestimated the cost of fasting on exam performance. However, the exemptions alleviated such bias in learning and made the treated students more willing to delay the fast for the exam. The results provide compelling evidence on how people distort valuable objective information and under-appreciate the cost of religious activities.

CHAPTER I

Improving College Choice for the Poorest Students: Results from Randomized Experiments in Centralized Admissions

1.1 Introduction

One of the most pressing issues facing K-12 and higher education is inequality in college access and success. Low-income students, facing various barriers at every stage of their educational pipeline, are persistently much less likely to attend college and particularly selective institutions (Holsinger and Jacob, 2009; Bailey and Dynarski, 2011; Li *et al.*, 2015). However, this income gap cannot be fully explained by the differences in academic preparation and financial affordability, given substantial policy efforts to increase K-12 school quality and financial aid for college students (see recent summaries in Glewwe and Muralidharan, 2016; Page and Scott-Clayton, 2016; Fryer, 2017). In recent years, the complex transition from high school to college has been increasingly recognized as an important barrier for students, particularly those from disadvantaged backgrounds (Lavecchia *et al.*, 2016; Page and Scott-Clayton, 2016).

College choice, whether and where to go to college, is one of the most crucial choices in life. Millions of high school graduates make their college choices every year all over the world, but their choices are often not made optimally due to informational and behavioral barriers. In particular, even when low-income students reach the college choice stage, they are more likely than their high-income peers to apply to and enroll at 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, 2017b). Undermatched college choice significantly lowers their chances of college and career success (Howell and Pender, 2016; Dillon and Smith, 2017a; Kang and Torres, 2018; Ovink *et al.*, 2018). The undermatch problem is prevalent not only among American low-income students, but also in many other countries such as Chile (Hastings *et al.*, 2018), China (Loyalka *et al.*, 2017), and Russia (Prakhov and Sergienko, 2017). New policy solutions are needed to effectively address the inequality in college access and match on a global scale.

In the past decade, behavioral interventions have been proposed and implemented as promising policy tools to help students navigate the complex transition from high school to college. 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), and Damgaard and Nielsen (2018). Despite the rapidly growing literature of college choice interventions, two key policy questions remain unanswered. First, most studies focus on the decentralized admissions systems in the United States and Canada. Little is known about what works in other contexts (Dinkelman and Martínez, 2014; Hastings *et al.*, 2018; Peter *et al.*, 2018). In particular, many countries use centralized college admissions with mandatory entrance exams and a simplified application process, in which information is not the main barrier but students need to make strategic college choice decisions.[1] Moreover, many developing countries may impose additional institutional barriers, which call into question the effectiveness of those interventions that have shown promise in the current literature as well as in policy efforts. Second, personalized advising and assistance may be more effective than low-touch information provision to improve

[1]In countries like Brazil, Chile, China, Germany, Greece, India, South Korea, Turkey, and the United Kingdom, college admissions operate through national exams and a centralized application and admission system. Many American colleges are starting to use the Common Application, a platform through which students may submit the same application to as many colleges and universities as they like.

college-going decisions, but not much is known about the successful scale-up approaches of these high-cost and often labor-intensive advising programs in both the decentralized and centralized systems.[2]

This chapter addresses the first question by examining college choice behaviors and admissions outcomes of low-income students in a centralized college admissions system.[3] I conducted large-scale randomized controlled trials (RCTs) with high school graduates of the 2016 graduation cohort in one of the poorest provinces in China. Collaborating with the local government and high schools in Ningxia province, I designed and implemented the *Bright Future of China Project* to understand and address the academic undermatch problem by improving students' college choice behaviors.[4] This large-scale experimental design helps to make progress on several important research questions: (1) What are the sources of undermatch regarding application behaviors in centralized college admissions? (2) Do behavioral interventions in college choice improve student access and match in centralized college admissions? (3) Can the framing of the interventions (e.g., delivery channels) amplify the treatment effects?

The analysis begins by documenting the full extent of student-college academic undermatch in centralized college admissions. Using administrative data on students' college admissions, the results indicate that - as is true in the U.S. decentralized admissions system - academic undermatch is also prevalent in centralized admissions in China. Undermatch is defined as a student enrolls at a college with lower peer quality, or she does not enroll at college. Using a very conservative bandwidth of 0.25 standard deviations, students from urban families have an average undermatch rate of 23.1 percent.[5] Rural students - a common

[2]Low-touch and low-cost informational interventions can be provided to a large number of students (e.g., Hoxby and Turner, 2013; Bergman *et al.*, 2019; Bird *et al.*, 2017), but may not be sufficiently effective. In contrast, personalized advising may be very effective, but may not be scalable (Bettinger *et al.*, 2012; Carrell and Sacerdote, 2017; Oreopoulos *et al.*, 2017; Oreopoulos and Petronijevic, 2018).

[3]I examine the potential solutions to the second question in the second chapter. Both chapters focus on a student's decisions about which colleges to apply to because the centralized college admission system simplifies the application process.

[4]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.

[5]Undermatch is defined to equal one if a student enrolls at a college with a median College Entrance

proxy for poor students in developing countries - are 9.8 percentage points more likely to undermatch. This poverty gap in undermatch persists when I compare two students in the same classroom with the same CEE score and demographic characteristics.

Next, a unique dataset of students' college applications is used to investigate the correlation between a set of college choice behaviors and admissions outcomes. Results show that many students do not use appropriate college application strategies that would improve their college access and match. Low-income students more often lack information and knowledge about college choice. The descriptive evidence indicates that the most important factor in college choice behaviors driving undermatch is that students do not use targeting strategies. Successful targeting strategies require a sound understanding of college admissions and intensive data analysis in order to apply to a mixed set of reach, peer/match, and safety colleges. Hoxby and Turner (2013) describe this as expert advice, which high-income students often use in their college choices.

The central question of interest is: *What behavioral interventions could improve college applications and admissions in a centralized system?* I differentiate the roles of information and knowledge/strategy in the process of college choice.[6] Access to information - college attributes, admissions policies, and admission data in prior years - is not enough because students must use knowledge and strategies to make thoughtful decisions, particularly in centralized systems where the online application process is simplified. Nevertheless, students lacking knowledge and guidance may have behavioral responses to the complicated decision-making. For example, they may just use very "simple and sometimes naïve" strategies, such as choosing a college or major based on its name.

To help students make behaviorally-informed college choices, I designed college appli-

Exam (hereafter CEE) score 0.25 standard deviations lower than her own CEE score, or if she does not enroll at college. The undermatch rate will increase when using a smaller bandwidth.

[6]A few interventions are considered to be very ineffective or not in need in centralized systems, such as reminders (students receive a series of text messages from the Department of Education), application fee waiver (students already pay for the very low college entrance exam testing and application fee), information/nudge/assistance of the application procedure (simple and straightforward), information about college return (almost all students are motivated to attend college), and information about college cost (centrally provided by the Department of Education).

cation assistance through a knowledge-based approach that promotes a student's learning and use of these knowledge and strategies. I prepared a comprehensive college choice and application guide - a "college application textbook" - to help students gather information and learn the rules and principles for college-going decisions. In this way a student can make an informed college choice by themselves. In particular, the focus of the guide was making accurate predictions of admission probability, which is the key to college choice in centralized systems. Using stratified cluster randomized experiments, I examined the effectiveness of the college application guide through different school-based delivery channels: (a) a guidebook, (b) a school workshop, and (c) one-on-one advising.[7] As summarized in Table 1.1, In the *Bright Future of China Project - Ningxia 2016*, 32,834 high school graduates in Ningxia were randomly assigned to one control group or three treatment groups that received a guidebook, a workshop, and/or advising during the five-day college application period in late June.

The experimental evidence shows that knowledge-based application assistance interventions, during a short time period with low costs, substantially improve students' college access and match outcomes by shaping their college application behaviors. The intensity/touch of the intervention increases its effectiveness, and the one-on-one advising seems to be the most effective intervention. All of the three interventions have increased students' use of targeting strategies, and have nudged students to consider out-of-province (and higher quality) college options. This result is consistent with the theoretical expectations and the focus in developing the interventions.

This chapter makes several contributions to the literature. First, it provides new evidence about student-college academic undermatch and its potential sources regarding application behaviors in centralized admissions. The results suggest that simplifying the college application process and providing a centralized information platform may not fully address the undermatch problem. Given the importance of application strategy and sophistication

[7] I proposed and examined these school-based channels because high school is the primary source of information and assistance for low-income students. Furthermore, making the researcher-initiated behavioral interventions into school-based practice, or public policy at large, would be a potentially effective approach to scale up these interventions and to improve school effectiveness.

in centralized admissions systems, the use of strategies - especially those for making admission probability predictions - is the key driver of matched college choices and admissions. I also find that distance is an important factor that shapes students' college choices, and that focusing on in-province colleges would limit students' high-quality college opportunities because low-income regions are often "college deserts" (Hillman, 2016; Hoxby, 2000; Long, 2004; Miller, 2017; Ovink *et al.*, 2018). I do not find evidence that preferences for tuitions, admissions quotas, and majors largely affect students' college access and match in centralized admissions.

More importantly, the results contribute 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). The existing literature primarily focuses on U.S. students. Very limited evidence is available about what works in improving college decisions in centralized systems or in developing countries.[8] Centralized admission is widespread across countries in both K-12 and higher education. While it streamlines and simplifies the application process, it may require strategies and sophistication in decision-making, so that one may expect to see differences in the effectiveness of existing behavioral interventions (Pathak and Sönmez, 2013; Chen and Kesten, 2017). This chapter provides novel evidence on the impacts of application assistance interventions in the use of knowledge, as opposed to information provision and application simplification, from the largest centralized college admissions market in the world.

The intervention designs build on many prominent approaches, including information provision (Hoxby and Turner, 2013; Goodman, 2016; Peter and Zambre, 2017; Herber, 2018), text message reminders (Castleman and Page, 2015), 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.*, 2019; Castleman and Goodman, 2018), and school workshops

[8] Many U.S. college admissions 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?"

or services (Oreopoulos and Ford, 2019; Bowman *et al.*, 2018). This chapter demonstrates an effective researcher-initiated, problem-solving intervention approach that focuses on the instruction and learning of new knowledge in college choice and application. The finding of the effectiveness of knowledge-based interventions such as guidebooks, workshops, and personalized advising, is also consistent with recent literature in decentralized or non-selective admissions. For example, Oreopoulos and Ford (2019) show that school workshops that guide students' college applications increase community college applications and enrollment. Carrell and Sacerdote (2017) suggest that college mentoring substitutes for the potentially expensive and often missing ingredient of skilled parental or teacher time and encouragement. However, the treatment effect does not derive from simple behavioral mistakes or a lack of easily obtained information.

The rest of this chapter is organized as follows. Section 2 lays out the theoretical framework of college choice, as well as the institutional background of Chinese college admissions and the Bright Future of China Project. Section 3 describes the full extent of undermatch in centralized admissions and sources of undermatch regarding college application behaviors, and shows the potential for behavioral intervention design. Section 4 presents the experimental designs and results. Section 5 offers discussions and policy implications. Section 6 concludes.

1.2 Background

1.2.1 Theoretical Framework: Information and Knowledge in College Choice

College choice is complicated. Following the economic literature on decision-making under uncertainty, college choice models (e.g., Manski and Wise, 1983; Kane, 1999; Long, 2004; Perna, 2006; Jacob *et al.*, 2018) assume that rational and forward-looking college applicants choose from a feasible set of colleges the one that maximizes their expected utility.[9] Early work concentrates on the human capital model, which states that the economic costs and benefits of college attendance are the most important factors affecting students' college choice behavior (see a summary in Perna, 2006). This approach is particularly useful when examining the effects of academic ability and financial resources, including family income, tuition, and financial aid on college choice. However, it is not sufficient for understanding other sources of observed differences across student groups, for example, differential information available to decision-makers (Manski, 1993).

The remaining differences in college choice that are not explained by the standard human capital theory are widely believed to result from two behavioral factors: (a) heterogeneous preferences including observed or unobserved "non-monetary benefits," risk aversion, and attention, and (b) unequal access to information. Drawing on both economic and sociological perspectives, Perna (2006) conceptualizes that four contextual layers shape a student's college choice decisions: the individual's habitus; school and community context; the higher education context; and/or the social, economic, and policy context. Given these intangible contextual layers, heterogeneous preferences result in different weighting of the benefits and costs of college attendance (Paulsen, 2001). Students may also have different expectations about benefits and costs because of unequal access to information (DesJardins and Toutkoushian, 2005), which is largely determined by their family and school social capital (Coleman, 1988). More recent literature has examined the impact of various individual preferences over college and major attributes, such as selectivity, college type, cost, distance, and consumption amenities (e.g., most recent studies include Jacob *et al.*, 2018; Ovink *et al.*, 2018),[10] as well as the lack of sufficient information and guidance on college choice behavior

[9]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 chapter focuses on a student's decisions about to which colleges to apply.

[10]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.

(see summaries in Page and Scott-Clayton, 2016 and Castleman, 2017).

Regarding college choice decision-making, this chapter differentiates the roles of information and knowledge, although people may use the two terms interchangeably. One approach defines information as structured data or simple facts, which subsequently becomes knowledge when information is put into a context and when it is can be used to make predictions and inferences (Tuomi, 1999). I refer to information as the individual ability (e.g., test scores) as well as college attributes and admissions policies available to a student at the time when she makes college choices. I define knowledge as the sophistication and strategies used to navigate the complicated decision-making process, which needs intensive instruction and learning. Therefore, college application assistance can be seen as a knowledge-based intervention rather than simple information provision.

For example, Hoxby and Avery (2013) note that the advice of experts concerning college application is to apply to several "peer" colleges, a few "reach" colleges, and a couple of "safety" colleges. This advice itself is simple information. However, knowing this advice is not enough because identifying college types requires knowledge and skills to gather and analyze the SAT/ACT and GPA information of each college. As discussed in Subsection 1.2.2, centralized college admissions require sophisticated and strategic applications more than decentralized systems. Applicants should compare both the potential utility received from attending each college and the probability of being admitted to each college (or the risks of being rejected). Differential access to information and knowledge (application assistance) would jointly determine college applications.

In this chapter, I expand the multi-stage college choice model (e.g., Hossler and Gallagher, 1987; DesJardins *et al.*, 2006) by detailing the steps to make informed decisions about which colleges to apply to. Students in either decentralized or centralized systems need to go through this decision-making process. I also classify the requirements for information and/or knowledge in each step. This theoretical framework helps to understand why students make undermatched college choices and how behavioral interventions would improve their decisions.

Students need to gather information about the set of available colleges and then make use of decision-making knowledge (e.g., admissions policies, prediction rules) in order to choose the final college application lists. This process includes six key steps:

(1) Decide whether or not to attend college;

(2) Search for college and major information (in Chinese college admissions, a student may have more than 20,000 college-major options);

(3) Collect and understand college admissions policies;

(4) Identify reach, match/peer, and safety colleges based on predictions of admission probabilities;

(5) Apply appropriately for a list of reach, match, and safety colleges (and majors);

(6) Explore alternative admissions and special programs such as early admissions and income-based affirmative action programs.

Steps (1)-(3) could easily be assisted by light-touch informational interventions, but steps (4)-(6) require intensive knowledge and sophistication. Inadequate knowledge and lack of application guidance may be a primary explanation for differences between students in their behavioral responses, even with similar access to information.[11]

The key to the six-step informed college choice is to correctly predict college admission probabilities in order to identify different college types. The ability to carefully and sophisticatedly make use of information is crucial for predictions, given the significant uncertainty in college admissions. Even with the best available *ex ante* information, college admission is still a two-sided, incomplete information problem (Chade *et al.*, 2014). Colleges would like to recruit the best candidates in the application pool but do not know students' true types or enrollment probability. When applying for college, students do not know their chances of

[11] 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; Netzer and Benkert, 2018), but it can also result from lack of knowledge and skills 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.

getting into each college (admission rate can be very low in top schools even for very high achieving students). In order to construct a thoughtful application list with a set of reach, peer/match and safety colleges, students have to make accurate predictions of admission probabilities using available information, and possess the knowledge and skills to make these inferences. Additionally, college choice is not always unconstrained in either decentralized or centralized systems. Costly applications (e.g., complex process and application fees) limits the number of applications. Admissions policies (e.g., the admission assignment mechanism) would also require students to be sophisticated in choosing their final college application lists.

The persistent inequality in college access and match is one of the key issues in the transition from K-12 school to college. College undermatch is largely attributable to the lack of (and/or errors in) information and knowledge for the six-step college choice, particularly among minority, low-income, and first-generation students. Conventional K-12 and higher education policies emphasize the adequacy of academic preparation and financial aid for disadvantaged students, but they are not sufficient. The growing importance of information and knowledge in shaping students' college choice behaviors calls for an ongoing understanding of the roles of information and knowledge, and how we could improve the existing behavioral interventions.

1.2.2 Centralized College Admissions

Every year, millions of high school graduates apply for college admissions worldwide. In some countries, like the United States and Japan, college admissions are decentralized. Students must apply separately to each college and colleges make their decisions independently.[12] Admissions results are often based on a number of performance measures including SAT or ACT scores (mandatory test improves college enrollment, see Bulman, 2015; Hyman,

[12]Many American colleges are starting to use the Common Application, a platform through which students may submit the same application to as many colleges and universities as they like. Many K-12 school admissions are centralized, such as in Amsterdam, Boston, Paris, and New York (Hafalir *et al.*, 2018), as well as in all Chinese cities.

2017), high school grades, recommendation letters, and personal statements. In many other countries like Brazil, Chile, China, Germany, Greece, India, South Korea, Turkey, and the United Kingdom, college admissions are operated through national exams and a centralized application and admission system. Admissions results are solely determined by the exam score and students' ordered college choices.

Both mechanisms have desirable and undesirable properties. In decentralized admissions, students can be matched with multiple colleges but applications are costly.[13] In centralized admissions, applications are simplified and costless because students only need to submit their rank-order lists of colleges; they can often apply to many colleges at a minimal or zero cost simultaneously. But additional barriers may arise: students may be admitted at most to one college, which imposes large risks and uncertainties when choosing which colleges to apply to.

The centralized college admissions market. Since the seminal paper by Gale and Shapley (1962) and the influential papers by Balinski and Sönmez (1999) and Abdulkadiroğlu and Sönmez (2003), the centralized admission mechanism - notably the Deferred Acceptance algorithm - has been an important policy innovation for school and college choice, which is considered to improve efficiency, welfare, and match.[14] Following the framework in Balinski and Sönmez (1999) and Arslan (2018), the student-college centralized matching market consists of a set of students and a set of colleges. Each college has a non-zero quota. Colleges rank students according to their entrance exam scores, the priority in the matching process. A college prefers a student over another student if and only if the first student has a higher priority. During the college application process, each student submits a rank-order

[13]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 colle, ge attendance and degree completion of free SAT reports.

[14]The school/college choice mechanism design literature is motivated by American K-12 school centralized mechanisms, rather than the college decentralized mechanisms (Abdulkadiroğlu *et al.*, 2005, Pathak and Sönmez, 2013, Agarwal and Somaini, 2018, Calsamglia *et al.*, 2018). Centralized clearinghouses have also been adopted in many markets, including college admissions in many countries (Machado and Szerman, 2017), and in the U.S. K-12 school choice.

list of colleges (and majors). A centralized student-college matching mechanism operates the assignment using students' rank-order lists and exam scores that determine students' priorities.[15]

Students derive utility from being admitted to a college. Students rank each college based on their individual preferences and tastes. The academic match literature, particularly in cultures where families emphasize college quality, assumes that students mainly rank colleges based on observed college quality rankings. Other preferences (e.g., tuition, location) can be seen as restrictions on the rank-order list. A distinctive feature of centralized admissions is that students could have an (accurate or inaccurate) idea of the matching outcomes (e.g., admissions cutoffs) for each college in the past years (Yenmez, 2018).[16]

Students in centralized systems apply to a rank-order list of colleges to maximize their expected utility. This utility maximization meets several constraints due to some common institutional barriers in centralized admissions. First, admissions are implemented with a list of restricted length in many real-world examples (Arslan, 2018; Chen and Kesten, 2017). For example, students in Ningxia Province studied in this chapter could only apply to four selective colleges and six majors within each college. Many students do not even apply to the complete list of colleges. Under the Deferred Acceptance mechanism, but with restrictions on the application, truthful revelation of preferences is no longer the optimal strategy. Second, many centralized matching systems only match subjects (e.g., students) to at most one option, which imposes a high risk of all applications being rejected. Students have to evaluate the *ex ante* admissions probability of each college based on the admissions outcomes (e.g., cutoff or median admission scores) in the prior years. Moreover, the centralized admissions may occur during a very short time period. In order to make a thorough search and

[15] In contrast, a decentralized matching mechanism may include many other factors besides standardized test performance, such as alumni or teacher rating, personal rating, and extracurricular rating. Since academic index does not account for everything in the admissions process, even the academically top-rated students cannot have a 100% admission probability at top schools. As for rank-order lists, early admissions or legacy admissions can be seen as a form of rank-order (over all other colleges in the regular application period).

[16] Decentralized systems could also have developed such data-based information systems to improve the student-college matching process and results.

assess college fit, students have to search for and analyze a large volume of college/major information from many reliable or unreliable sources.

Uncertainty plays a central role in the college choice and application strategy, particularly in centralized admissions (DesJardins and Toutkoushian, 2005; Chade *et al.*, 2014). The strategy for each student is a function of her test score (known), preferences and valuations for each college (based on complete or incomplete information), predicted *ex ante* admission probabilities, and other individual idiosyncratic factors. To make an optimal decision that maximizes her expected utility, a student encounters a search cost and a learning cost. Search cost has been intensively addressed through information provision and the simplification of the application process in prior research as well as in policy practice (e.g., centralized clearinghouse; in the Chinese context, college admissions results in the prior years and information about tuition, location, and quota are publicly provided to students by the Department of Education). However, the learning cost of predicting *ex ante* admission probabilities (or assessing uncertainties) could be substantially higher in centralized systems than that in decentralized systems. Students may only gain an accurate prediction of their admission probability based on a sound understanding of the admissions mechanism and the admissions outcomes in prior years. Both of these require deep learning and use of the knowledge in college-going decision making. For students who lack such knowledge and skills, intensive advising and assistance, rather than information provision, may be the only effective intervention to improve their college choices and applications.

1.2.3 Context: College Choice in China

China has a centralized college admissions system. China's college application and admission procedures are centralized at the province (state) level. Students only compete for college-major spots with peer applicants within the same province and STEM/non-STEM track. The process begins with the administration of the annual national College Entrance Examination (hereafter and before CEE) in early June, similar to the SAT/ACT in the

United States. The CEE scores are the sole criteria used to rank the priorities of students in college admissions. High school seniors take four subjects: Mathematics, Chinese, English, and track composite. Students choose either the STEM track with exams in physics, chemistry, and biology; or the non-STEM track with exams in history, social studies, and geography.

Next, all the Chinese colleges allocate their college-major admissions quotas to each province and publish their quota and tuition information. All the information is provided to students by the provincial Department of Education in mid-June. After knowing their CEE scores in late June, students submit their college application lists to the provincial Department of Education. Students who choose not to apply to any college do not submit applications. The list includes 4-10 colleges for each institutional tier, varying across provinces, in which they rank colleges and majors within each college (4-6 majors for each college). The submission process is very simple in that students only need to type in the college and major IDs in the online system.[17]

College application and admission proceeds by institutional tiers. Tier 1 includes the nation's elite colleges. Tier 2 and Tier 3 consist of non-elite public and private four-year colleges, respectively. Tier 4 includes three-year vocational colleges, which resemble community colleges in the U.S. 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. Each year in each province, about 20 percent of CEE takers are not admitted by a college: About 10 percent do not apply for any colleges, and the other 10 percent apply but are rejected.

Student application eligibility is limited to colleges in certain tiers based on their CEE score and the tier-specific admission cutoff scores, which are determined by the total number of spots and the distribution of the CEE score within the province. That is, students with CEE scores above the Tier 1 cutoff are allowed to apply to colleges in all tiers. Students

[17] In Subsection 1.11.1, I show and explain a typical college application form in China.

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. Within each tier, there might be additional special admissions 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. A student can apply to more than 50 colleges (she does not have to) or just 1 college.[18]

Based on their CEE scores, each student is then matched with only one college-major in their college application list through a predetermined matching mechanism. Like many Chinese provinces, Ningxia currently uses a parallel mechanism like the Deferred Acceptance mechanism (see discussions in 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. Some students may not be accepted by any college. If a student declines the offer, or does not receive an offer, she must wait until the following year to retake the CEE and participate in the matching process again. The alternative is to enter the job market with a high school degree, or to enroll at a university outside China.[19]

This centralized system has several features that change or call into question the effectiveness of those interventions that have shown promise in decentralized systems such as that in the United States. If so, then different or redesigned interventions may be necessary for countries with centralized admissions systems. Centralized admissions could induce both positive and negative effects on the undermatch problem for disadvantaged students. Therefore, the combined effect is theoretically ambiguous.

On the one hand, as discussed in the U.S. literature, disadvantaged students undermatch due to both preference heterogeneity and the complexity of the application process.

[18] It is not an optimal strategy to apply to too many or too few colleges. For the former, special admission programs, which are undermatched options for some students, complete admissions before the normal programs. Applying for these special programs without careful consideration may result in undermatched college admissions. For the latter, applying to too few colleges limits one's chances of trying reach or match colleges, and increases the risk of all applications being rejected.

[19] Very few students make the oversea college plan 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.

The undermatch problem should be minimized (a) in countries where students have homogeneously strong preferences for college quality, (b) in centralized systems where the college application process is simplified with "one exam for all," and (c) where college admissions depend exclusively on exam scores and applications without sending scores, institution-specific essays, or reference letters, which are common requirements in the college-level decentralized admissions systems.

On the other hand, some other institutional features may result in behavioral barriers to low-income students relative to their high-income peers. First, after students are informed about their CEE scores, they need to submit their applications in about three to five days. This time requirement calls into question their ability to quickly gather and analyze a large volume of college and major information. Second and importantly, students in China have to strategically "game" the risky college admissions with accurate predictions accompanied by uncertainty. They are allowed to list a limited number of colleges, typically four to ten colleges in each tier. Students have to apply for a mixed set of reach, peer, and safety colleges to maximize their opportunities of being admitted by a reach or peer college, and to minimize the risk of being rejected by all of the colleges to which they have applied.

Moreover, students have to choose colleges and majors simultaneously. The match process is college-then-major, which imposes additional barriers in college applications (Bordon and Fu, 2015). A student would be considered for admission to colleges in her rank order, and will be reserved for admission to the first college for which her CEE score qualifies. After that, the college will make the final admissions decision based on the student's major choices and the major rank order. If all of the majors on the student's list have higher admissions cutoffs than her score, she will be assigned to some other major outside her application list by the college - but only if she indicates her willingness to being admitted to a major (with a lower major admissions cutoff) she has not applied to. If the student does not qualify for all of the majors she has applied to, and if she denies a flexible major assignment, she will be rejected by the college. She will not be considered for admission to any other colleges in the same institutional tier and will move down to lower tiers. Therefore, to minimize such risks, it is also important for students to pick reach, peer and safety majors within each college they apply to. This process emphasizes the roles of both information and individualized assistance with sufficient knowledge and skills to make an informed decision.

1.2.4 The Bright Future of China Project

In a previous research project with coauthors (Loyalka *et al.*, 2017), we found that low-income students in the Chinese centralized college admissions substantially undermatch in their college choices and admissions.[20] Upon finishing that project, and motivated by the emerging research and policy efforts to support low-income students in college access and match in the United States, I worked with a research team at the China Center for Education and Human Resources Research at Peking University to start the *Bright Future of China Project*. The project is to bring together both our research team's decade-long intensive experience of advising college choices (knowledge and strategies) and the use of large-scale data to provide effective college-going interventions. Our ultimate goal is to inform policy and contribute to the college choice literature by altering the college and major choice behavior of low-income students through various cost-effective behavioral policy interventions.

With an emphasis on low-income students, I have collaborated with local governments to conduct experimental research in several of the most underdeveloped provinces in China. This chapter examines the first-year interventions in 2016 in Ningxia province, one of the poorest provinces with a typical centralized college admissions system (Appendix Figure 1.5 presents the project timeline). Ningxia, officially the Ningxia Hui Autonomous Region, has the third smallest total GDP in China where Muslims are more than 38% of its population. Most of the region is desert, making Ningxia one of the poorest provinces in northwestern China. Appendix Figure 1.6 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:

[20] Using novel data on college applications of a more recent cohort of high school graduates (the control group in the main experimental sample in 2016), Subsection 1.3.1 provides an extended analysis.

$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. I focus on both rural and urban students because, in fact, most of them are from low-income families and lack information and knowledge in college choices and applications. Each year, nearly all high school graduates (about 60,000) take the College Entrance Exam.[21] About 90% of the exam takers apply to college and 85% are admitted to college, but fewer than 10% are admitted to elite colleges.

1.2.5 Data Sources

I use unique, large-scale student level administrative data for the universe of 2016 high school graduation cohorts in Ningxia province. The data are provided by the Ningxia Department of Education and the Ningxia Education Examination Board, the provincial centralized administration office of the College Entrance Exam and college admissions. Using accurate administrative data to analyze the entire population of applicants in a college matching market (not a sample), and without problems of missing data and sample attribution, I am able to identify students' college application behavior (strategies and preferences), admissions outcomes, and enrollment decisions. I can also map students to their exact high school classes and neighborhoods to examine school and neighborhood effects.

The confidential student-level data consist of four separate sources: (1) College Entrance Exam Registration Data that include student demographic information (e.g., gender, race, urbanicity, age, repeater identifiers, track, high school attendance records, school and class identifiers); (2) College Entrance Exam Score Data that include all the CEE score information (total score, and scores by subject), which I use to create track-specific standardized scores (high-stakes) and high school graduation test scores (low-stakes); (3) College Applications Data that include all the rank-order application lists that students submit to the Ningxia

[21]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.

Education Examination Board (college and major IDs in each rank order); and (4) College Admissions Data that include the final admissions results of all students who have submitted their applications (college and major IDs, and the relevant college-major information). I linked all of these student-level administrative data using a unique student identifier. This data linking was done in several rounds in the secure data room at the Ningxia Department of Education. The analytical data are de-identified.

I merged the college-major level information (address, tuition, quota, prior-year admissions scores) with the student-level data to study their college choice strategies and preferences. This college-major level information is also provided to students during the college application period by the Ningxia Department of Education in print books.[22] I merged in additional college-major data (e.g., elite college identifiers, national college ranking). Combing student-level data and college-major data, I also created a number of variables measuring college match outcomes and college application behaviors, which will be discussed in detail in Subsection 1.3.3 and subsection 1.11.1. Finally, I extracted the track-tier specific admissions cutoff scores from the Ningxia Department of Education's official website. In addition, confidential school finance data from the China Ministry of Education (access provided through the Institute of Economics of Education at Peking University) were used for the school-level randomization in 2016.

1.3 The Policy Challenge

1.3.1 College Undermatch in Centralized Admissions

During the past decade, student-college academic undermatch has drawn concern from education researchers and policymakers. It is widely believed that approximately 20 to 70 percent of American high school graduates undermatch (estimates vary across data, sam-

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

ple, and methods). Researchers have used various definitions of undermatch mainly due to data availability or specific research questions (see summary discussions in Rodriguez, 2015; House, 2017). The early pioneering work uses a categorical matrix (e.g., selectivity categories) that defines undermatched students as those who enroll in a college selectivity that is below the college selectivity to which they had access based on their academic credentials (Roderick *et al.*, 2008; Bowen *et al.*, 2009; Smith *et al.*, 2013). The problem is that this method assumes no higher education institutional enrollment constraints. Instead, Dillon and Smith (2017b) use the difference in student ability percentile and enrollment size weighted college quality percentile to measure the match between students and colleges.

Hoxby and Avery (2013) propose an alternative definition of college match by comparing a student's SAT/ACT score with a college's incoming freshman cohort median score. They define a safety school such that the college's median score is 5 to 15 percentiles below the student's own, and a peer school when the absolute value of the difference between the college's median test score and the student's own is within five percentiles. One appealing feature of this measure is that it corresponds more closely to the probability of admissions than the categorical matrix measures. House (2017) uses one sample of recent high school graduates from Tennessee to compare the different definitions of undermatch. She concludes that using different definitions results in large variation in undermatch rates.

To best describe the extent of undermatch in the Chinese centralized college admissions system, I take advantage of the centralized system itself. Moreover, the data enable me to study an entire college matching market to construct several college access and match measures. Speaking to the literature, I first constructed an "undermatch" indicator, which equals to 1 if a student is admitted to a college with a peer median CEE score 0.25 standard deviations lower than her own CEE score, or when the student is not admitted to any colleges. The choice of 0.25 s.d. as a conservative threshold is based on the practical experience of college choice advising in China. Table 1.2 shows that this measure presents a slightly smaller overall undermatch rate (28.6%) to that (33.5%) from the five-percentile threshold as proposed in Hoxby and Avery (2013).[23]

Centralized college admissions provide a valid identification condition for estimating the impacts of college choices on college admissions. Admissions outcomes in decentralized systems are not only determined by demand-side preferences and application behavior, but also by supply-side factors such as geographic priorities (as an admissions policy), capacity constraints, and academic screens (Corcoran *et al.*, 2018). This results in omitted variable bias in using student-level data to identify how college choices affects admissions (Jacob *et al.*, 2018). In contrast, centralized admissions are solely determined by exam scores and the college choices of students. Holding exam scores constant, I can directly identify the relative undermatch between two groups (e.g., rural vs. urban, female vs. male) by comparing the differences in college admissions outcomes. This approach expands the details of college access and match in the analysis on both the extensive margin and the intensive margin.

The extensive margin measure is a dichotomous indicator of being admitted to one college (=1). For the intensive margins, I consider several measures in addition to the undermatch indicator as described above, which jointly denote the college match results. I use college median, mean, and minimum scores of the incoming freshmen in the same year to measure contemporaneous college match.[24] Holding one's CEE score constant, a negative difference in college peer quality - resulting from the same year's college admissions - means that a student has "wasted" her CEE score to be undermatched with that college. To minimize the potential bias of using the college admissions results in a single year and in a single province to denote college quality, and to compare results across years, I use two national college quality measures: standardized score and ranking percentile.[25] Using these

[23]Table 1.8 shows that the results remain qualitatively unchanged using other thresholds. Throughout the chapter, all the results using these undermatch indicators remain consistent.

[24]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 s.d. as their "college median/mean score." In Figure 1.7, I show that the results are very consistent using different measures for this group of students. The estimates correspond to the regression results of identifying the rural-urban gap in Table 1.3.

[25]Using college admissions 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 is now published at siminedu.com to assist all Chines high school graduates in their college choices.

five intensive margin measures of college match, I construct a single index using principal component factor analysis as the primary college match outcome measure.

Quantifying undermatch. In this section, I use the sample of untreated students from the 2016 graduation cohort in Ningxia. I included both the control group of the randomization sample and those not in the randomization sample that will be introduced in section 1.4.[26] Table 1.2 documents the full extent of academic undermatch in a typical centralized admissions system in China. The rows represent students' CEE score quartiles, indicating to which quality level students have access. The columns show the college quality level to which a student is eventually admitted. The student-college match in centralized admissions is very similar to that in decentralized admissions, notably in the U.S. literature (e.g., Smith *et al.*, 2013, Dillon and Smith, 2017b). Students show an assortative matching pattern such that 65.9 percent of students concentrate along the diagonal. However, about 25 percent of students are admitted to a college that is one quality level below the level to which they have access. The change in overmatch is not accordingly symmetric as noted by Dillon and Smith (2017b): 9.1 percent of students end up with overmatched colleges based on the quartile matrix. Using the primary undermatch indicator as discussed above, there are 28.63 percent of students who are admitted to a college with a median CEE score 0.25 standard deviation lower than their own CEE scores.[27] About 12% of the students enrolled at overmatched colleges.[28]

The undermatch statistics vary substantially by student achievement levels. The third CEE quartile students, facing the choice between four-year colleges and three-year vocational colleges, have the highest undermatch rate. The pattern is different from that in decentralized systems. First, students in the highest CEE score quartile have a lower undermatch rate than lower-achieving students. More than 90 percent of students in the highest CEE score quartile

[26]The college access and match measures are constructed using the whole cohort data.

[27]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.

[28]The few students who were in the low CEE quartiles but enrolled at elite colleges were mainly through special programs.

are admitted to the highest quality quartile colleges. This is partly due to the tier-specific admissions policy that guides higher-achieving students to apply for higher-quality colleges, while lower-achieving students have less information about differentiating college quality. Second, not being admitted to and then not enrolling at any college is an important source of undermatch not only for the lowest CEE score quartile - similar to U.S. students who only have access to two-year colleges - but also for students in the second and third quartiles. This result suggests that college admissions are pretty risky in centralized admissions.

1.3.2 Poverty Gap in College Undermatch

The undermatch literature in decentralized admissions has well documented that disadvantaged students are more likely to undermatch. I focus on the poverty gap in college match, using rural *hukou* as a proxy for poverty. *Hukou* is household registration, and is the primary source of income inequality in China.[29] In developing countries, there are often large gaps between rural and urban families in socioeconomic status, parental education and income, and information. Figure 1.1 provides graphic evidence of the sizable rural-urban gap in student-college undermatch. Compared with students in urban families, rural students are much more likely to undermatch. This gap exists among both high-achieving and low-achieving students, and it is larger among higher achieving students.

I use a linear model to formally estimate the poverty gap in college access and match:

$$Y_i = \beta_0 + \beta_1 * Rural_i + \gamma * X_i + \varepsilon_i \tag{1.1}$$

where β_1 measures the rural-urban gap in the college match outcomes Y_i, with standard errors clustered at high schools. I add additional covariates X_i (CEE score, demographics, and class fixed effects) stepwise to examine how well these factors explain the observed

[29] The urban-rural income gap has been institutionalized by the *hukou* system since 1955 in China, under which all households had to be registered in the locale where they resided and also were categorized as either "rural" or "urban" households. Assigned at birth on the basis of the mother's registration status, *hukou* limits rural residents from migrating into the urban areas and entitles few of the rights and benefits that the government confers on urban residents, such as permanent employment, medical insurance, housing, pensions, and educational opportunities for children. See more discussions in Wu and Treiman (2007).

rural-urban gap.

Table 1.3 presents results on the rural-urban gap in college access and match. Within each panel that uses different model specifications, each cell shows the estimates from a separate regression. Column (1) of Panel A shows that urban students have an average admission rate of 87.9 percent, but rural students are 6.5 percentage points ($p<0.01$) less likely to be admitted by a college.[30] This gap persists even when I control for CEE scores and demographics. Panel B adds the CEE scores to control for differences in academic achievement. Holding CEE scores constant, the results show that different college choices and applications largely contribute to the rural-urban gap of college undermatch. In Panel C, I further add student-level demographics - gender, race, repeater, STEM track, and age - to further control for heterogeneous college choice preferences. The large rural-urban gap persists. In Panel D, I additionally control for class fixed effects (an average class has about 59% rural students). Students in the same class within the same high school may receive information and assistance from teachers simultaneously, which reduces about one third of the rural-urban gap in college admissions. Because most students attended high school in their hometown county, controlling for neighborhood effects does not change the results once I control for school or class effects.

Columns (2)-(4) of Table 1.3 present very consistent results on the intensive margins. Rural students are admitted to colleges with about 0.267 standard deviations lower quality in the single college quality index (column 2).[31] Even comparing with urban classmates in the same high school class who have the same CEE scores and demographic characteristics, rural students are admitted to colleges with about 0.1 standard deviations lower quality. Column (3) excludes students who are not admitted to college and shows that, even conditional on the selected sample of those who gain admissions, there are still sizable and statistically

[30]This is due to both a lower application rate and a lower admission rate conditional on application. Table 1.10 shows that rural students are 2.9 percentage points less likely to apply to college.

[31]Table 1.9 decomposes the five outcomes that I use to construct the college match index. Results are very similar using the in-province measures or the national measures. Table 1.10 shows that rural students are less likely to apply to and enroll at college, they are more likely to retake CEE in the next year, and they are less likely to enroll at matched or overmatched colleges.

significant gaps between rural and urban students. Using a large threshold of 0.25 standard deviations, column (4) shows that rural students are much more likely to undermatch.

The results confirm that, as is true in the U.S. decentralized admissions system, academic undermatch is pervasive in centralized systems as well. Moreover, disadvantaged students in centralized systems statistically significantly and substantially undermatch more than advantaged students.[32] However, how to improve college access and match in these centralized systems remains an open and challenging question. As is evident in Table 1.3, the poverty gap between rural and urban students largely persists within the same classroom where they share the same information and assistance from teachers. This calls for innovative policy interventions to more effectively improve college access and match.

1.3.3 The Potential for Behavioral Interventions at Scale

I explore student behaviors in college applications in order to set the stage for designing policy interventions to improve college access and match at scale. I use the full college applications data of those untreated students in 2016 to test what matters in college choices. Based on the theoretical framework of the six key steps in college applications, I use the observed behaviors in college applications to construct a list of key strategies and preferences that students have. Appendix subsection 1.11.1 provides a detailed description of these measures.

First, I focus on three sets of measures of application strategies in centralized admissions, which include (1) general advice, (2) targeting, and (3) special programs. The targeting strategy is the core of the knowledge in college choice and application, which enables students to apply to a targeted set of reach, peer/match, and safety colleges. It requires intensive knowledge and sophistication (e.g., understanding the policies and utilizing historical data) to make the appropriate predictions and decisions.[33]

[32] In the regression results of Table 1.3, female, minority, older, and lower-achieving students are more likely to undermatch. Repeaters are much less likely to undermatch, suggesting the potential benefits of repeating grade 12 in information, knowledge, and/or experience of college choices and applications.

[33] Many students do not understand the underlying mechanism of college admissions: Only rank matters,

Next, the other important aspect of college choice is preference. However, students' preferences and tastes are individual-specific and strictly unobservable. Particularly in constrained college applications, revealed preferences may not be exactly students' true preferences. I constructed three sets of proxy preferences using the applications data. The first set includes college tuition and quota. Low-income students may prefer low-tuition colleges, and risk-averse students may prefer colleges with larger admissions quotas (Hoxby and Avery, 2013; Loyalka *et al.*, 2017). The second set includes geographic location (indicators for colleges out of province or in the economically developed regions). Distance is one important factor shaping a student's college choices, but focusing on in-province colleges would limit other high-quality college opportunities (Hoxby, 2000; Long, 2004; Hillman, 2016; Miller, 2017; Ovink *et al.*, 2018).[34] The last set includes a few major group indicators (e.g., the most popular ones such as economics, computer science, and the least popular ones such as agricultural-related majors).

Figure 1.2 presents a set of pairwise correlations comparing the primary college access and match outcome - college median CEE score - with the (partial) list of college choice strategies and preferences as described above. The set of targeting strategies seem to be strongly (and statistically significantly) correlated with improved college access and match. The general advice is also positively correlated with college match. Applying to more out-of-province colleges is associated with increased college median scores.

In Table 1.4, I extend Equation 1.1 by including the measures of the constructed strategies and preferences to examine how much they could explain the rural-urban gap in college

not raw scores. They naively compare their CEE score in this year with raw college admissions scores, which results in large errors when identifying college types. Figure 1.8 shows the distribution of student applications. The X-axis shows the distance of college median score and a student's own score. I separately present the distributions for a student's first choice and fourth (last) choice in the match tier. It clearly shows that, though correctly centered, some students apply to match colleges (and in the order that the first choice should aim higher than the last one). However, a large proportion of students apply to colleges that they will be substantially undermtached to, or apply to colleges that they have a nearly zero chance of getting into. In 2018, I provided online advising to more than 30,000 high school graduates across China (not a randomized experimental sample). Score equating and targeting the match colleges is the single most important question that the students had.

[34]High-quality colleges concentrate in the economically developed regions in China. Ningxia province, as a low-income region, lacks high quality colleges.

match. Consistent with the differences in admissions outcomes, rural students are less likely to use the appropriate strategies such as flexible major assignment and targeting, and are much less likely to choose out-of-province colleges.[35] Comparing the changes in the coefficients on rural-urban gap, targeting strategies explain the largest variation in the outcomes among all the strategies and preferences, even controlling for CEE score, demographics, and high school fixed effects. Alternatively, an Oaxaca-Blinder decomposition tells a very consistent story. About 0.078 s.d. of the 0.132 s.d. (59%) estimated rural-urban gap (rural mean: 0.058; urban mean: 0.189) in the college match index, controlling for CEE score and demographics, is explained by rural-urban differences in the college choice strategies and preferences that I constructed. The set of targeting strategies explain a 0.065 s.d., or 83% of the explained difference in the rural-urban gap.

Together, this section's descriptive analysis demonstrates the potential for behavioral interventions to improve college access and match by affecting the college choices and application strategies of students. Overall, students in low-income areas such as Ningxia show that they may have informational and behavioral barriers in college applications. They generally do not use appropriate strategies that would increase their college access and match. Poorer students show more severe problems and therefore are more likely to undermatch. Among all the strategies and preferences, the targeting strategies - which require intensive data analysis based on knowledge and understanding of college admissions - seem to be the most important element of a behavioral intervention for low-income students. I now turn to intervention designs and then randomized experimental designs to examine the effectiveness of these interventions.

[35]In the U.S. decentralized college system, Dillon and Smith (2017b) find that numbers of applications and the average distances to schools applied to both increase with parental wealth and education.

1.4 Experimental Design

1.4.1 Interventions: Promoting the Learning of Knowledge in College Choice

In the first year (2016) of the *Bright Future of China Project*, I explored effective college-going interventions for poorly informed students in centralized college admissions. The testable hypothesis is that students lack information and knowledge in college choice, but it is unknown what works to help students gain such expertise. Considering what actions an expert counselor or a very sophisticated student would take for the decision-making, I prepared a comprehensive college application guide. I designed different school-based channels for the delivery of the application guide, including (1) a print guidebook, (2) a school workshop, and (3) individualized advising. These channels differ in the intensity of instruction/touch.

In the theoretical framework, I categorize what a student needs in college choices and applications on two levels: (1) college information - including cost, return, curriculum, major - and admissions policies, and (2) knowledge that can be used for strategic college choice and application. Besides access to accurate information, students need to learn a variety of knowledge and skills to navigate the complex, yet risky, process. They need to learn not only about the college and major options available to them, but also sophisticated decision-making strategies. If information is the key behavioral barrier, the guidebook could have similar effects with the other two more intensive interventions. If knowledge (strategy and sophistication) matters in centralized systems, I would expect the intervention effects to be larger because of more intensive "human instruction" in the school workshops and personalized advising.

The intervention design builds on the Application Strategies approach of the Expanding College Opportunities project in Hoxby and Turner (2013).[36] It combines features of both

[36] I do not incorporate the other interventions in Hoxby and Turner (2013) including net cost, application fee waiver, and parent intervention. In the Chinese centralized admissions, students are provided with tuition

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 focus exclusively on the instruction and learning of the specific knowledge in college choice during a very short time period (one week) when students apply for college. College choice may be the first life decision that a young adolescent would make, and they probably are not taught to make decisions. Therefore, I draw on the principle of "teach a man to fish" rather than "give him a customized fish." These transferable decision-making skills can be applied to other choices in life, in the hope of potential impacts from the interventions on longer-term college and life outcomes.

I led an army of experts that included professors and graduate students in the field of both K-12 and higher education policy, school counselors, and college admissions officers. I prepared the college application guide in the form of a guidebook. Using our expertise in advising college choice for more than a decade in China, and conducting additional learning from many prominent sources,[37] our research team produced a very comprehensive guidebook. With regard to the key steps and strategies in college choices and applications, the guidebook is designed to consist of four main "course" modules: (1) searching for college information, (2) understanding admissions policies, (3) equating CEE scores and identifying college types, and (4) applying to match colleges. subsection 1.11.2 provides more descriptions and sample pictures (Figure 1.10) of the guidebook.

To supplement the main modules, I also make use of both large scale (and confidential) databases as well as reliable information about colleges and college applications. In advising students about how to search for information, I provide a table that maps a list

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.

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

of recommended websites of college and major information (panel A of Figure 1.11).[38] To assist students with major choice, I use 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 the employment rate trends (panel B of Figure 1.11). Lastly, I provide detailed explanations of the college admissions policies (e.g., the Deferred Acceptance mechanism) and the corresponding applications strategies and tips.

Intervention 1: Guidebook. I provided guidance to students in the form of *a college application guidebook*. Students were told 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 is expected to be fully scalable and could be easily modified under different admissions policies and contexts across the country. It not only helps students gather information, but also facilitates their learning of the rules and principles necessary to make a knowledgeable decision for themselves. Similar to the behavioral interventions in the U.S., informing and advising students about college-going decisions is a public good that could potentially be integrated into the existing K-12 and higher education systems. Thus, an important policy question is whether the *college application guide* intervention works. If so, in what ways?

Learning knowledge and skills in the college choice process may be as complex as the process itself. It is not surprising that a guidebook alone may not well guide a student's self-directed learning. Following the traditional textbook-teacher instruction, I provided two additional human-intensive interventions: school workshops and one-on-one advising. Both are based on the guidebook, but use different delivery channels to provide deeper learning and personalized counseling opportunities to help students learn and understand the contents in the guidebook. These two interventions are similar to class lectures and after-school coaching in coursework.

[38]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.

Intervention 2: School workshop. With the assistance of the Ningxia Department of Education, I worked with local districts and high school leaders to plan and run the school workshops. To minimize the quality variations in the workshops, I selected a group of very knowledgeable experts - the editors of the guidebook - to give the workshops, using the same slides and scripts at each school. Workshops were announced one month ahead of time in the name of a joint research team from Peking University and Ningxia University. Each workshop lasted three hours and was moderated by a high-level school administrator. Figure 1.12 shows sample pictures of the workshops, during which all the four "course" modules were covered in detail by the speakers and a Q&A session was included.

Intervention 3: Individualized advising. I provided one-on-one advising invitations and contact information to students at the end of each workshop. The individualized advising was implemented through the two largest and popular online chat Apps in China (Wechat and QQ, similar to iMessage) using text, picture, and video. Students who self-selected to sign up for assistance were then assigned by the administrative assistants to one of the six core advisors. During the personalized advising, students were given personalized, detailed guidance and suggestions on specific colleges and majors, as well as text message reminders about application tips and deadlines as well. Figure 1.14 shows the real conversations between one treated student and one advisor.

The interventions - guidebook, workshop, and advising - are designed to collaborate with high schools for two reasons. First, in low-income regions and for disadvantaged families, high schools are a student's primary and most trusted information source. If students need to show up at a central location to receive the intensive instruction, schools should be the best place for these students. Second, the interventions as a whole are not very costly so reforming the existing education systems by integrating these interventions would be a promising approach to scale up such a public service, and thus to increase K-12 school effectiveness in preparing students for college access, choice, and success.

1.4.2 School Level Randomization

I conducted a cluster randomized controlled trial to evaluate the effectiveness of the proposed three interventions: guidebook, workshop, and individualized advising. 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 field 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.[39] 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.[40]

Implementation and take-up. Of the total 32,834 high school graduates in 2016 in 31 public high schools, 11,408 students were in 12 control schools. 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. School administrators distributed them to individual students when they came back in school to receive their score report.[41] Another 8,603 students in 7 schools were provided with both the guidebook and the workshop (T2).[42] Workshops were held during 22-24 June 2016, when students started to submit their college applications (the deadline was 27 June). All students in the workshop schools were informed one month ahead of time, and their parents were also invited to participate. During the workshop, I randomly provided students the opportunity for individualized one-on-one advising in five high schools (T3). The advising continued until students completed their applications.

[39] 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.

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

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

[42] 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.

Nevertheless, I was not able to identify an accurate take-up of the interventions because schools failed to track the actual "treated" students (guidebook receipts and workshop attendees), due to the lack of incentive and organizational capacity in these high schools.[43] I do have complete records of the conversations with students who received the individualized advising. According to the follow-up short surveys and field observations by our research team members, nearly all students received the guidebook. A few exceptions include those who were sick or out of province, and who had already decided not to apply. About 30 to 50 percent of students attended the workshop.

Summary statistics and validity. Table 1.11 provides summary statistics on the whole graduation cohort and the study sample, suggesting that the experimental sample is representative of the entire student population in Ningxia. The experimental sample has a 0.11 s.d. higher average CEE score, a 6% lower minority student fraction, and a 7% higher rural student fraction. About 60% of students are from rural families, about 30% are minorities (mostly Muslims), and about 20% of college applicants have repeated the 12th grade at least once. The average college admissions rate is 84 percent. Students on average attend college with lower-achieving peers than themselves, which is consistent with the prevalent undermatch phenomenon.

Mean student characteristics differ slightly between groups because the randomization used school-level finance data in 2013. T1 has more rural students, and relatedly, on average lower-achieving students. 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 1.5) and for students (for both the whole sample and the high-achieving sample in 2016; see Table 1.12).

[43] I actually provided monetary incentives to school administrators, and they were also recommended (but not required) to keep records by the district superintendents.

1.4.3 Econometrics

To estimate the intent-to-treat effects (ITT) of the guidebook and workshop, I estimate the following linear regression using OLS:[44]

$$Y_{ij} = \beta_0 + \beta_1 * T1(guidebook)_j + \beta_2 * T2(workshop)_j + X_i * \gamma + \delta_j + \varepsilon_{ij} \qquad (1.2)$$

where Y_i is the outcome of interest for student i in school j. $T1_j$ and $T2_j$ are indicator variables for school j receiving the guidebook treatment and the guidebook-workshop combined treatment respectively. T2 also includes a small proportion of students who received one-on-one advising (T3). δ_s are strata fixed effects. X_i includes a set of student characteristics, particularly a student's CEE score, to identify the "college choice" effect. I also control for demographics (gender, race, age, STEM/Non-STEM track, repeater) to account for group differences in college preferences. Given the school-level cluster randomization, including the school-level controls does not alter the estimates; standard errors are also qualitatively unchanged. All standard errors are clustered at the school level.

I first examine the admissions outcomes as described in Subsection 1.3.2, including both the extensive and intensive margins, because the interventions are primarily designed to improve college access and match.[45] Next, I examine a list of exploratory measures of college choice behaviors, including both strategies and preferences as discussed in Subsection 1.3.3. Following the undermatch literature, I examine a key heterogeneity in the treatment effect between high-achieving students eligible for selective college admissions based on their CEE scores, and their lower-achieving peers.

Since the take-up of T1 is more than twice that of T2 according to anecdotal evidence, the treatment-on-the-treated (TOT) effects would be of policy interest as well. I can use a Wald estimator to rescale the ITT effects by the take-up probabilities. The approximate

[44] Identifying the treatment effects of individualized advising is difficult and complex given students' self-selection. I provide detailed discussions in subsection 1.11.4.

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

TOT estimates provide a sense of the results if we could fully scale-up the interventions; for instance, making the workshop (and the associated learning of the "college application textbook") a mandatory part of the high school curriculum.

I address multiple hypothesis testing in several ways. I construct the admissions match outcomes within the same family of domains. These measures provide a complete picture of college match and are also highly correlated with each other. I also aggregate both the admissions outcome and choice behavior measures into several single indexes. Additionally, I apply the method proposed by List *et al.* (2016) to confirm the robustness of results.

A final 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 use randomization inference to assess whether the observed treatment effects are likely to have been observed by chance even if treatment had no effect (Heß, 2017). I report p-values from 1,000 permutations.

1.5 Results

1.5.1 Effects on Admissions Outcomes

Average effects of the whole experimental sample. Table 1.6 presents the main results of the intent-to-treat effects of the guidebook and workshop interventions on college access and match outcomes. Panel A shows the results from the main specification without school covariates, aggregated from student data as shown in Panel B of Table 1.5. Panel B includes school covariates. Each column of each panel reports coefficient estimates from a separate OLS regression of Equation 1.2, as well as the control group mean and standard deviation of each outcome for reference.[46] Randomization inference is based on permutation tests using 1,000 simulations. I report the randomization inference p-values in parentheses.

I find that both the guidebook and the guidebook-workshop combined interventions

[46]Results are similar when I do not control for student demographic covariates.

improve college access and match. On the extensive margin, column (1) of Panel A shows that offering guidebooks or school workshops causes students to be 2-3 percentage points more likely to be admitted to college, although imprecisely estimated due to the small number of clusters. In Table 1.13, I show that the interventions insignificantly increase college application by about 1 percentage point, likely through information provision and nudging. Comparing the two estimates, the interventions have increased the college admission rate - conditional on application - by about 1-2 percentage points.

Results on the intensive margin of college match show that treated students are admitted to statistically significantly and substantially higher quality colleges. Column (2) shows that students who are offered, and potentially read the "How to apply for college?" guidebook, are admitted to a college with a 0.094 s.d. ($p<0.001$) higher quality using the single college match index. This suggests a large and precisely estimated impact. If students remain unchanged in their college application behaviors, they would have to score about 0.09 s.d. higher on the College Entrance Exam to be able 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 period, and at a much lower cost than other K-12 education policies and interventions.

The ITT effects of the guidebook-workshop combined intervention (T2) are very similar. Treated students, on average, are admitted to colleges with a 0.076 s.d. ($p<0.05$) 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 (T2) are two or three times larger than guidebook alone (T1). For example, using the Wald estimator, the rescaled TOT effects on college match index range approximately from 0.15 s.d. to 0.23 s.d. The results confirm that the instruction and learning of college application knowledge is effective at helping students make better decisions in college choice. Comparing with the rural-urban gap as reported in column (4) of Table 1.3 (-0.176 s.d.), the fully implemented guidebook-workshop combined intervention could close this poverty gap in college match.

To check the robustness of defining college quality for those non-admitted, column (3) excludes students who are not admitted to college, and shows a smaller impact of the interventions. However, given that the interventions affect college admissions, this result is downward biased. Column (4) shows that treated students are 3-4 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 quite stable using either within-province or national measures.

In Table 1.13, 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.[47] The interventions also increase the share of students that are admitted to match/peer and overmatch/reach colleges. In Panel B of both Table 1.6 and Table 1.13, I control for school-level aggregated differences in students' CEE scores and demographics to account for potential bias from the observed, but statistically insignificant, differences between randomized groups. The results are close (slightly larger), which confirms the robustness of results.

Effects on high achieving students. Table 1.14 repeats the analysis for high achieving students who are eligible for selective college admissions based on their CEE score. Nearly all high achieving students, who are also highly motivated for college, apply to and are admitted by a college. Thus the interventions have 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 percent. The reason for this difference is that some high achieving American students may not apply to any college. Chinese students do apply for college. However, they may not know how to apply to the appropriate set of colleges due to lack of knowledge about

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

college choice strategies. Consistently, I find clear evidence that both the guidebook and the guidebook-workshop combined interventions have statistically significant impacts on college match for high achieving students in both the single index and itemized measures. Being offered and potentially reading the guidebook increases college match index by 0.058 s.d. ($p<0.05$), holding CEE scores and demographics equal. Being offered both the guidebook and workshop increases college match index by 0.08 s.d. ($p<0.001$).

Heterogeneity in the treatment effects. Figure 1.3 summarizes the heterogeneous effects. For the guidebook only intervention, the ITT effects are slightly larger for rural, female and minority students. For the guidebook-workshop combined intervention, female and non-minority students benefit more. These differences may result from differential take-up between groups. Lastly, the interventions do not have large impacts on repeaters. This is consistent with the information/knowledge theoretical framework. Repeaters already have at least one year's experience of searching for and using college choice information/knowledge. They are also highly motivated to improve their college admissions outcomes, more than those non-repeaters who have no sense of the college application process. Therefore, repeaters are more experienced and skilled in searching for and using the relevant information and knowledge.[48] Figure 1.9 shows similar results among high achieving students. One exception is that high-achieving repeaters also benefit from the workshop, particularly from learning the strategies and skills to make more accurate predictions.

Note on the effects of one-on-one advising. The one-on-one advising program was not randomized at the student level since it was in development phase. Students who attended one of the five school workshops (randomly chosen from the total seven T2 schools) were provided opportunities (vouchers) to receive individualized advising from experts on our research team. Table 1.19 shows that the take-up of the advising program is about 1.5 percent in the whole sample (N=119). About 3 percent of high achieving students who

[48] There were several "lottery winners" who were randomized into the one-on-one intervention in both 2016 and 2017. They reported many fewer questions in the second year. Survey data covering more repeaters support this explanation as well.

received the voucher eventually participated in the advising program (N=72).[49]

In the absence of individual-level randomization, it is difficult to estimate the causal impact of the one-on-one advising. In subsection 1.11.4, I use the random assignment of the 5 schools with vouchers and a Wald estimator to approximately quantify the individualized advising program effects. I find that the advising program increases admissions by 13.1 percentage points, admitted college peer median scores by 0.706 s.d., admitted college quality by 0.974 s.d. and 11 rank percentiles, and the college access and match index by 0.210 s.d.[50] This large effect is supported by the observed admissions outcomes of the treated students. Figure 1.4 presents the distribution of college peer median scores between students who actually received one-on-one advising and students in the control group. The figure clearly shows that, with few exceptions - particularly students who did not apply to any college because they chose to retake the CEE the next year - most students who received the individualized advising were admitted to their match or even reach colleges. They were not all shifted to the top of the outcome distribution because the intervention did not focus on the single task of increasing admitted college quality. Instead, during the personalized advising, students' multidimensional preferences were fully considered. One example is the preference for specific majors. If students are admitted to a reach college, they are more likely to be assigned to (lower quality) majors for which they have low interest. It is up to the preference of individual students to choose to be "a big fish in a small pond" or "a small fish in a big pond."

1.5.2 Effects on College Application Behaviors

In developing the *Bright Future of China Project*, I have identified a set of knowledge and strategies based on the structured six-step college choice decisions. I then presented

[49]More than 800 hundred users contacted us, but the screening process, based on school and student IDs, largely decreased the eventual take-up.

[50]The effects for lower-achieving students are smaller (and even negative), which are imprecisely estimated due to the small number of treated students. This is mainly because a few lower-achieving students finally chose to retake the CEE in the next year in order to go to higher quality colleges.

all the relevant knowledge in a guidebook, in a workshop, or in individualized advising. I expected that students would change their college application behaviors and make better decisions after learning and acquiring the knowledge and strategies. Based on the (partial) strategy and preference measures that I constructed from student application data as discussed in Subsection 1.3.3 (in-depth descriptions in Appendix subsection 1.11.1), I now test whether the improved college admissions outcomes are, in fact, from the changed application behaviors.

The expectation is correct. The patterns of college application behavior in Table 1.7 suggest that the interventions have substantially impacted students' strategies and preferences, and thus have improved their admissions outcomes. Column (1) uses the single principal-component factor index to summarize the effects on college application behavior. The guidebook-workshop intervention statistically significantly and substantially improves college applications. The guidebook alone also improves applications, in an imprecise and smaller magnitude. This result suggests that the guidebook-workshop combined approach is more effective than the guidebook alone approach in promoting student learning of the college choice and application knowledge.

Columns (2)-(7) test the differences between strategy and preference groups. Consistent with the descriptive results in Table 1.4, that the targeting strategies and location preference are the most important factors driving college match, the interventions have statistically significantly and substantially improved students' use of targeting strategies, and have shifted students from in-province colleges to out-of-province colleges, especially in the economically developed regions. In Table 1.15, I consider each individual item and find that I have helped more students to be able to identify college types accurately and apply for a combination of reach, peer, and safety colleges in a descending rank order. I find a similar intervention impact for high-achieving students. As shown in Table 1.16, providing both guidebook and workshop improves their college applications, similarly in targeting strategies and out-of-province college choices. The impact of the guidebook only intervention is smaller and statistically insignificant.

One important finding, consistent with the intervention design, is that the improvement in college access and match is not at the cost of substantially changing students' other preferences. Treated students appear to choose colleges with smaller admission quotas. But the impact on special programs and major preferences (except for an increase in the probability of applying for computer science majors) are mixed. One suggestive reason is that students may have strong prior preferences and motivated beliefs.[51] The general advice does not affect student applications.

I should note that the college application behavior characterized in the data is incomplete. Although I constructed five itemized measures in the targeting strategies, I do not fully characterize how students could make their optimal choices, given that they could apply to up to more than 50 colleges, and then 300 college-major options. Strategies and preferences are interrelated, so that students need to carefully consider all of them to construct their college and major application lists. Results in this subsection are very consistent with the expectations in designing the project, as well as numerous fieldwork observations and feedback not captured in the data.

1.6 Summary

In this chapter, using administrative data of college applications and admissions from one of the poorest provinces in China, I document that the student-college academic undermatch is prevalent in centralized college admissions. I find descriptive evidence that the undermatched college choices occur because students do not appropriately use college application strategies, especially the strategies that make accurate predictions of college admission probabilities.

I conducted large-scale randomized experiments to examine what works in improving

[51] What I do not show in the admissions outcomes is that, consistently and mechanically, treated students (in guidebook or workshop interventions) are more likely to be admitted to out-of-province colleges with smaller quotas. But the impacts on other preferences, such as the choice of a major, are statistically zero.

college choices and admissions for low-income students. The first-year program of the *Bright Future of China Project* provides evidence that the knowledge-based human instruction approach to college application assistance is effective in improving college access and match for low-income students. Comparing different delivery channels or instructional methods, I find that the intensity of the "human instruction" increases the intervention effectiveness. Personalized advising seems to be the most effective program, and it is also suitable for accommodating individual preference heterogeneity in college choice.

This chapter provides the proof of the effectiveness of behaviorally-designed college choice and application interventions, which emphasize the instruction and learning of the problem-solving knowledge and skills in college choice and application. The results build a stage for improving relevant policy designs, particularly scaling up the effective personalized advising. In addition, I have learned substantial new lessons that I could use to fine-tune the intervention designs.

First, even the poorest students report that the application fee and access to the Internet is not a major barrier for their college choice and application, even though access to the Internet is a common problem in both developing and developed countries (Dettling *et al.*, 2018). Rather, poor students lack the guidance to use information and knowledge in making their college-going decisions. Nearly all the students who reach the stage of taking the CEE are highly motivated to apply to college. But they do not know how to appropriately apply to college because they have not been taught about decision-making during the entire K-12 period, particularly making use of data analytics.

Second, the *ex ante* expectation was that the prediction and targeting strategies are the key components of the knowledge in college choice and application. This is particularly true in the centralized admissions systems, where students should consider the expected utility and the admission probability of each college. The experimental evidence in this chapter shows that providing instruction and assistance largely improves the use of these strategies. However, these data-based prediction and targeting strategies require intensive learning and data analysis, which limits the potential of scale-up. Other strategies (e.g., general advise) and preferences (e.g., out-of-province options) could be potentially assisted using the similar guidebook/workshop approaches in the manner of information provision, which can be done by the centralized systems or schools. But a student still has to make accurate predictions of college admission probabilities and appropriately use the targeting strategies, which is similar to what the personalized advising program provides to the treated students. We need innovative scalable policy solutions to simplify the instruction and learning process.

Third, school leadership matters. In low-income regions of both developed and developing countries worldwide, where parents generally have very low schooling, high schools and teachers are the primary sources of information and assistance for students. But school leaders and teachers lack the knowledge and incentives to provide effective guidance. For example, one school failed to plan and run the workshop because the administrators thought that the workshop was not a required task in their accountability assessment. Many teachers attended the workshops as well and asked lots of "simple sometimes naïve" questions. Ningxia has switched from sequential admissions (Boston) to parallel admissions (Deferred Acceptance) since 2010, but many teachers still taught their students the wrong strategies that were based on the Boston mechanism.

This chapter makes the first-step attempt to improve college access and match by shaping student college choice behaviors among low-income students in centralized admissions. Results suggest that intensive personalized advising is the most promising policy lever. Targeted information in the personalized advising is human labor intensive, and would be difficult to implement at scale in practice as opposed to simple, general information (Corcoran *et al.*, 2018). An important policy question is how to scale up the personalized advising programs, which require a sufficient number of expert advisers. I address this question in the follow-up study in 2017, as summarized in Chapter II.

1.7 Main Figures

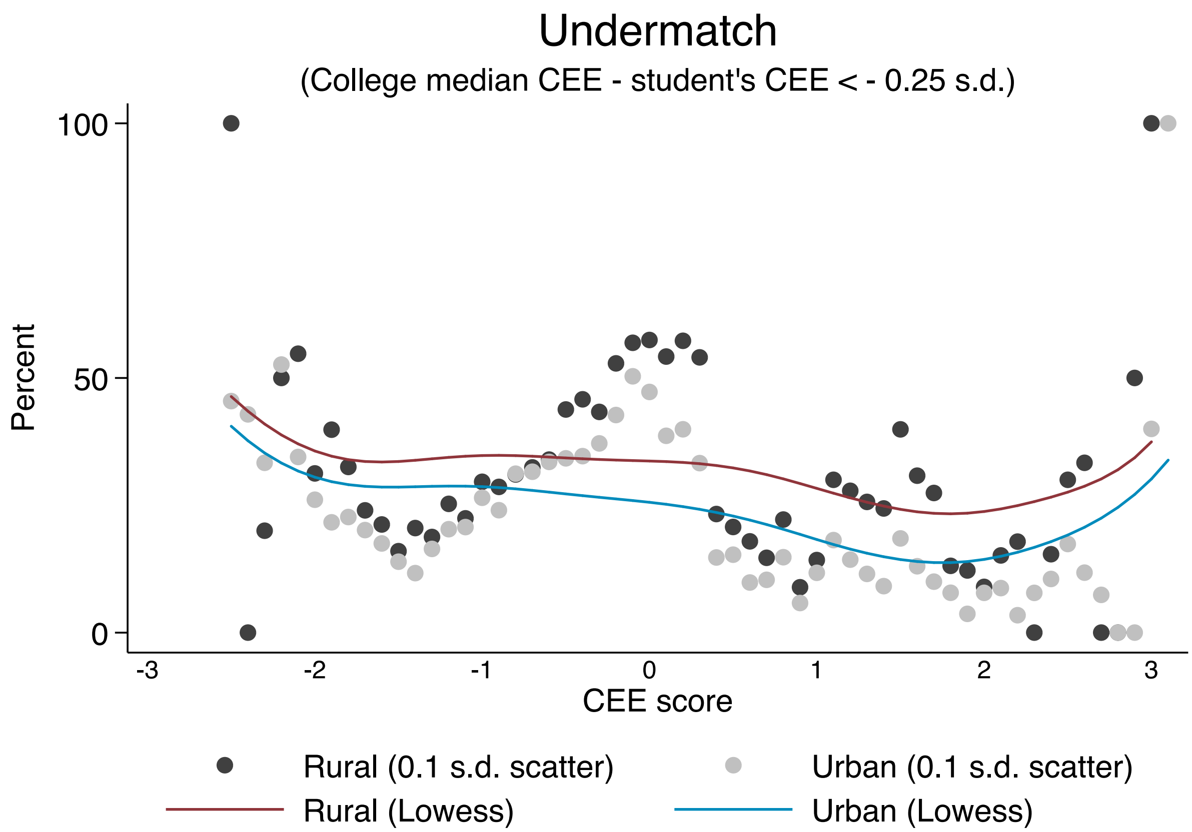


Figure 1.1. Rural-urban gap in academic undermatch

Notes: This figure plots locally weighted regression lines of academic undermatch rates between rural and urban students on the x-axis of students' CEE score (standardized). Each dot represents the average undermatch rate for a 0.1 s.d. bin. Undermatch is defined as a student being admitted by a college with median CEE score 0.25 s.d. lower than her own CEE score or not being admitted by any college. The sample includes 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.

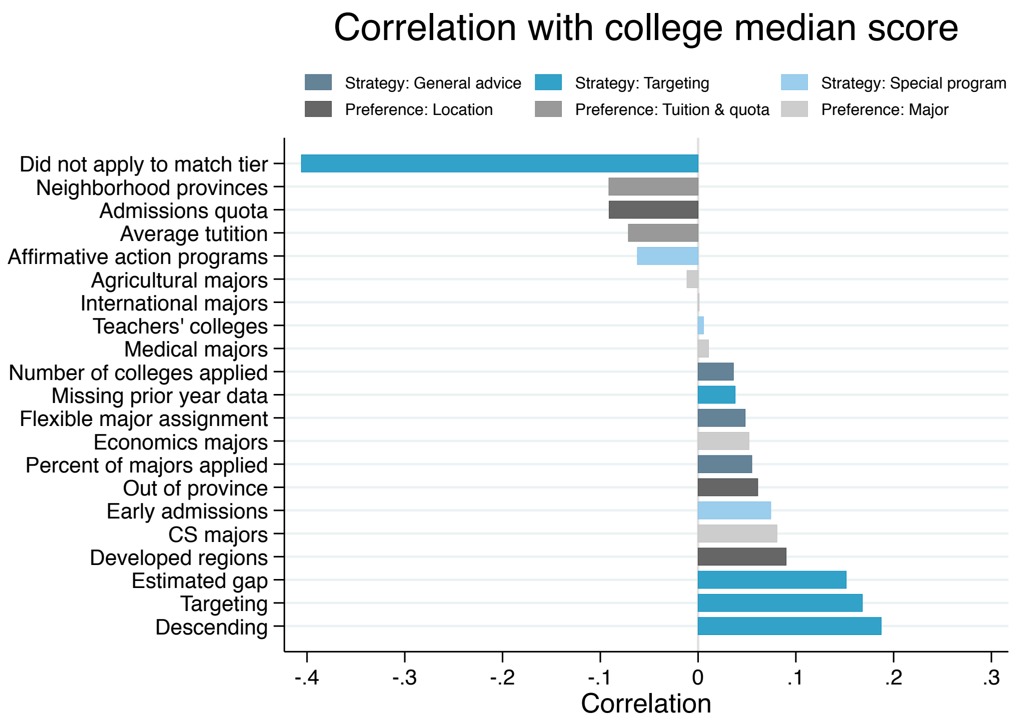
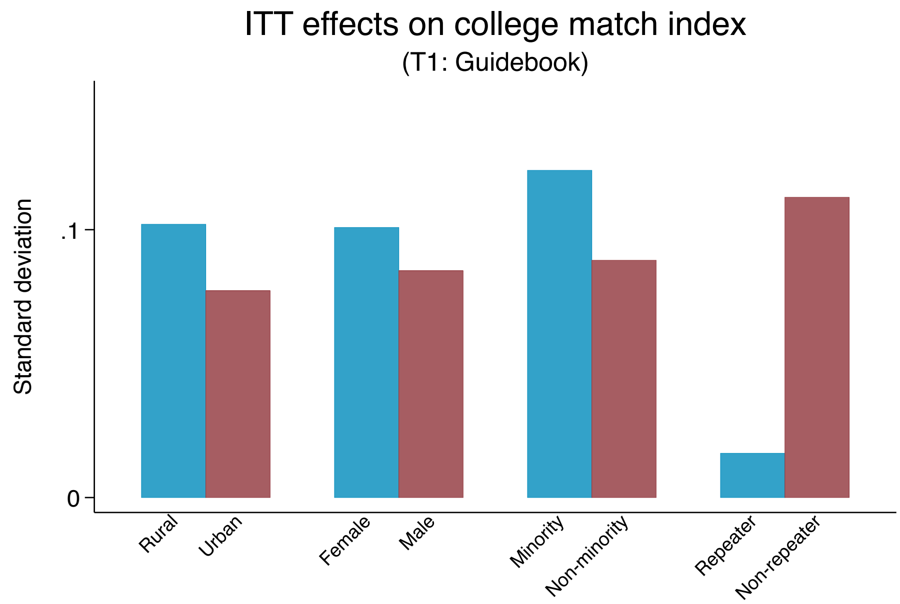
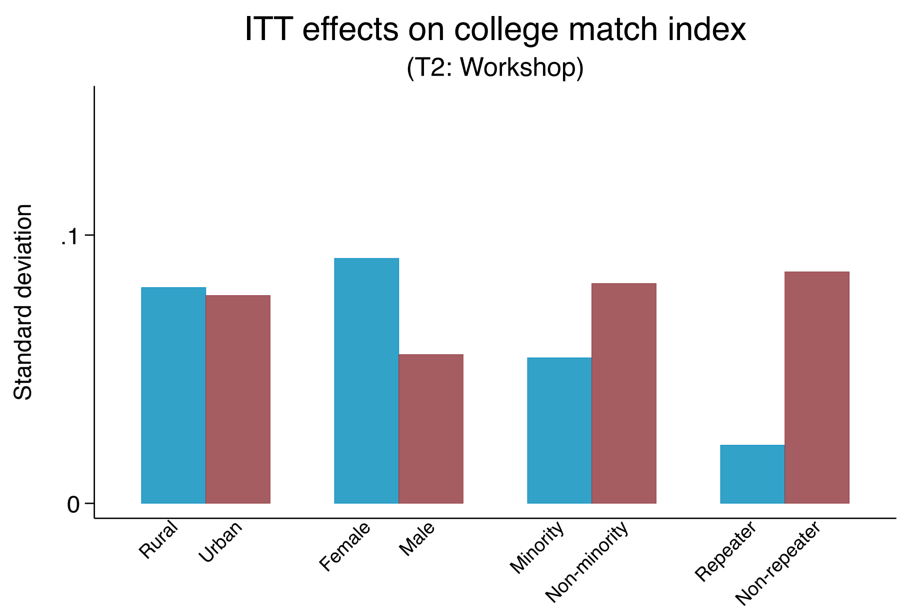


Figure 1.2. Correlates of college choices and admissions outcomes

Notes: This figure plots pairwise correlations between college match outcome (college median CEE score after adjusting for one's own CEE score and demographics) and several individual-level college choices and applications characteristics (strategies and preferences), as described in Appendix subsection 1.11.1.



(a) T1: Guidebook



(b) T2: Workshop (based on guidebook)

Figure 1.3. Heterogeneity in the ITT effects: All students

Notes: This figure plots heterogeneous ITT effects among the whole sample of the interventions on college median score from the OLS regression Equation 1.2, but with each subsample (e.g., rural students vs. urban students) separately.

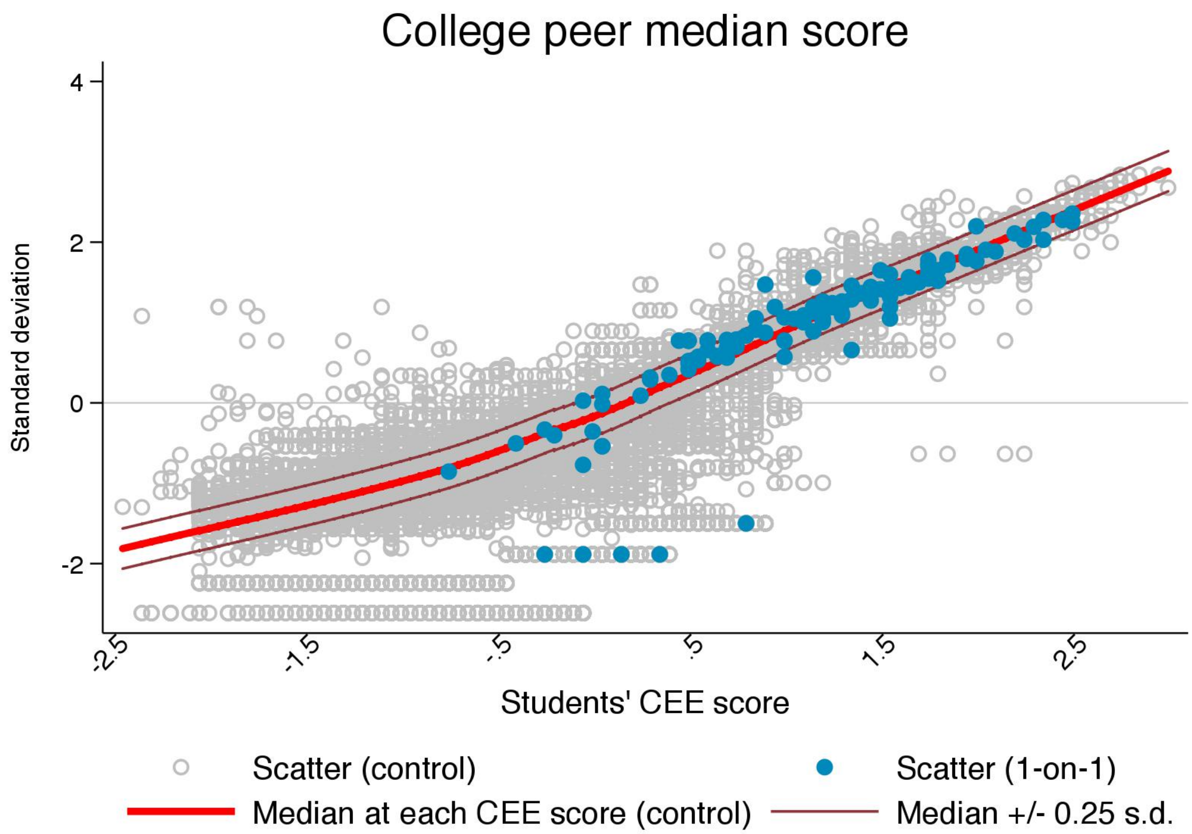


Figure 1.4. Distributions of college peer median score between students who received 1-on-1 advising and students in the control group in 2016

Notes: This figure plots the distribution of college peer median score between students who actually received 1-on-1 advising and students in the control group in 2016. For better illustration, each scatter summarizes students with a bin of 0.05 s.d. CEE score. The red line shows a locally weighted average of the outcome, and the maroon lines show the lower and upper borders of the match college range (0.25 s.d.). The bottom scatters indicate the tier-track specific lowest scores for those who are not admitted to a college.

1.8 Main Tables

Table 1.1. Experimental design: Interventions in 2016

	Interventions in 2016				
Question	What works in improving college access and match?				
Sample	All 32,834 students in 31 public high schools in three cities				
Randomization	School level within strata				
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	30-50%	ITT
Treatment 3	1-on-1 advising	5	6,025	1.5%	TOT

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. 1-on-1 advising in 2016 was not randomized at student level that we provided access to advising to students who attended the workshop in 5 of the 7 school workshops. Take-up rates for guidebook and school workshop in 2016 are from anecdotal evidence (school survey and field observations).

Table 1.2. Extent of academic undermatch: College access vs. college choice

	College quality quartiles (enrolled in) (N=31,777)					Percent Undermatch		Percent Overmatch
CEE quartiles (access to)	1st Quartile (Highest)	2nd	3rd	4th Quartile (Lowest)	No college	(0.25 s.d.)	(5 pctl)	(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 percentiles and 0.25 standard deviations as cutoffs, respectively. **Undermatch** is when a student's own CEE score is 0.25 standard deviations (or 5 rank percentiles) higher than here 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 deviations lower than her admitted college's median CEE score.

Table 1.3. Poverty gap in college access and match

	Admission (=1) (1)	Index (s.d.) (2)	Index* (s.d.) (3)	Undermatch (=1) (4)
Urban mean	0.879	0.173	0.153	0.231
	[0.326]	[1.008]	[1.032]	[0.422]
A. No controls				
Rural-urban gap	-0.065***	-0.267**	-0.215	0.098***
(β_1)	(0.015)	(0.134)	(0.169)	(0.016)
B. Control for CEE score				
Rural-urban gap	-0.055***	-0.161***	-0.088***	0.092***
(β_1)	(0.010)	(0.025)	(0.018)	(0.011)
C. Control for CEE score and demographics				
Rural-urban gap	-0.067***	-0.176***	-0.079***	0.097***
(β_1)	(0.011)	(0.026)	(0.014)	(0.012)
D. Control for CEE score, demographics, and class fixed effects				
Rural-urban gap	-0.043***	-0.099***	-0.034***	0.058***
(β_1)	(0.008)	(0.015)	(0.008)	(0.008)
N	31,777	31,777	26,776	31,777

Notes: This table reports the OLS regression (Equation 1.1) results of the rural-urban gap in various college access and match outcomes, using the universe of the untreated sample in 2016. The top two rows report mean and standard deviation (in square brackets) of each outcome variable for urban students. Each cell in Panels A-D is from a separate regression, showing the coefficient on the indicator of "rural *hukou* residence" (β_1). Column (3) only includes students who were admitted to a college. **Student-level demographics** include indicators of female, minority, repeater, and STEM track, and age by June 7, 2016 (=1 if at least 18 years old). Standard errors in parentheses are clustered at high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 1.4. College choices and the poverty gap in admissions outcomes

		Outcome: Index of college match						
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rural-urban gap		-0.176***	-0.123***	-0.074***	-0.069***	-0.035***	-0.035***	-0.031***
(β_1)		(0.026)	(0.018)	(0.012)	(0.012)	(0.009)	(0.009)	(0.009)
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)
School FE		No	No	Yes	Yes	Yes	Yes	Yes
Observations		31,777	28,806	28,806	28,806	28,806	28,806	28,806
R-squared		0.589	0.657	0.665	0.666	0.706	0.706	0.710

Notes: This table reports the OLS regression (Equation 1.1) 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. Application behaviors are constructed using the full applications data, as described in Appendix subsection 1.11.1. Column (1) shows the rural-urban gap in college admissions using the full untreated sample (same as in column (3) of Panel C in Table 1.3). Column (2) replicates the same analysis using the applicant sample. Column (3) controls for high school fixed effects (results are similar using class fixed effects). Columns (4)-(7) add the strategy and preference measures (principal component factor indices) stepwise. All regressions include a student's CEE score and other demographic covariates. Standard errors in parentheses are clustered at high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Estimates from regressions without school fixed effects in columns (4)-(7) are -0.113, -0.067, -0.067, and -0.054, all statistically significant at 1%.

Table 1.5. Balance checks: 2016 RCT

	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)			

Table 1.6. ITT effects on college access and match outcomes: Guidebook and workshop in 2016

	Main outcomes				Outcomes in Index (column 2)				
	Admission	Index	Index*	Undermatch	College median	College mean	College min	Quality	Ranking
	(=1)	(s.d.)	(s.d.)	(=1)	(s.d.)	(s.d.)	(s.d.)	(s.d.)	(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 1.2) results of the ITT effects of the guidebook and workshop interventions in 2016 on a family of college access and match outcomes. **Admission** denotes whether a student was admitted to college. **Index** measures college match, using principal component factor analysis based the five continuous outcomes in columns (5)-(9). **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/min scores** are constructed using all the admissions data in Ningxia in 2016. **Quality (standardized)** measures college quality using national data on college (admissions scores, inputs and employment data) from 1996-2017, and **Ranking** is the corresponding ranking percentile. All regressions control for student-level covariates (CEE score and demographics) and strata fixed effects. School covariates are aggregated mean values of student CEE score and demographics. 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 1.7. 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]
			A. Without school covariates				
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)
			B. With school covariates				
T1 (guidebook)	0.091*	0.066	0.127**	-0.062	0.004	0.115**	-0.002
	(0.082)	(0.360)	(0.019)	(0.355)	(0.930)	(0.036)	(0.970)
T2 (workshop)	0.099*	-0.01	0.083	0.086	-0.019	0.137**	0.015
	(0.078)	-0.912	(0.179)	(0.296)	(0.654)	(0.016)	(0.768)
N	29,591	29,591	29,591	29,591	29,591	29,591	29,591

Notes: This table reports the OLS regression (Equation 1.2) 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 1.11.1. All regressions control for student-level covariates (CEE score and demographics) and strata fixed effects. School covariates are aggregated mean values of student CEE score and demographics. 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%.

1.9 Additional Figures

The Bright Future of China Project (Ningxia) Timeline

2015/10	Proposal to Ningxia Department of Education
2016/03	Approval of the proposal
2016/04	Data collection and randomization
2016/06	*College Entrance Exam*
2016/06	Interventions (during *college applications*): Guidebook, workshop, 1-on-1
2016/08	*College admissions*
2017/01	Data collection
2017/02	New proposal approval
2017/05	Teacher survey
2017/06	*College Entrance Exam*
2017/06	Interventions (during *college applications*): Teacher incentives, 1-on-1 using ML
2017/08	*College admissions*
2018/01	*Data collection*

Ongoing...

Figure 1.5. Project timeline

Notes: This figure plots the general timeline of the Bright Future of China Project (Ningxia). Interventions in both 2016 and 2017 were primarily implemented during the short college application periods. Individualized advising was also provided in early August when students applied to non-selective colleges.

Figure 1.6. Location of Ningxia

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

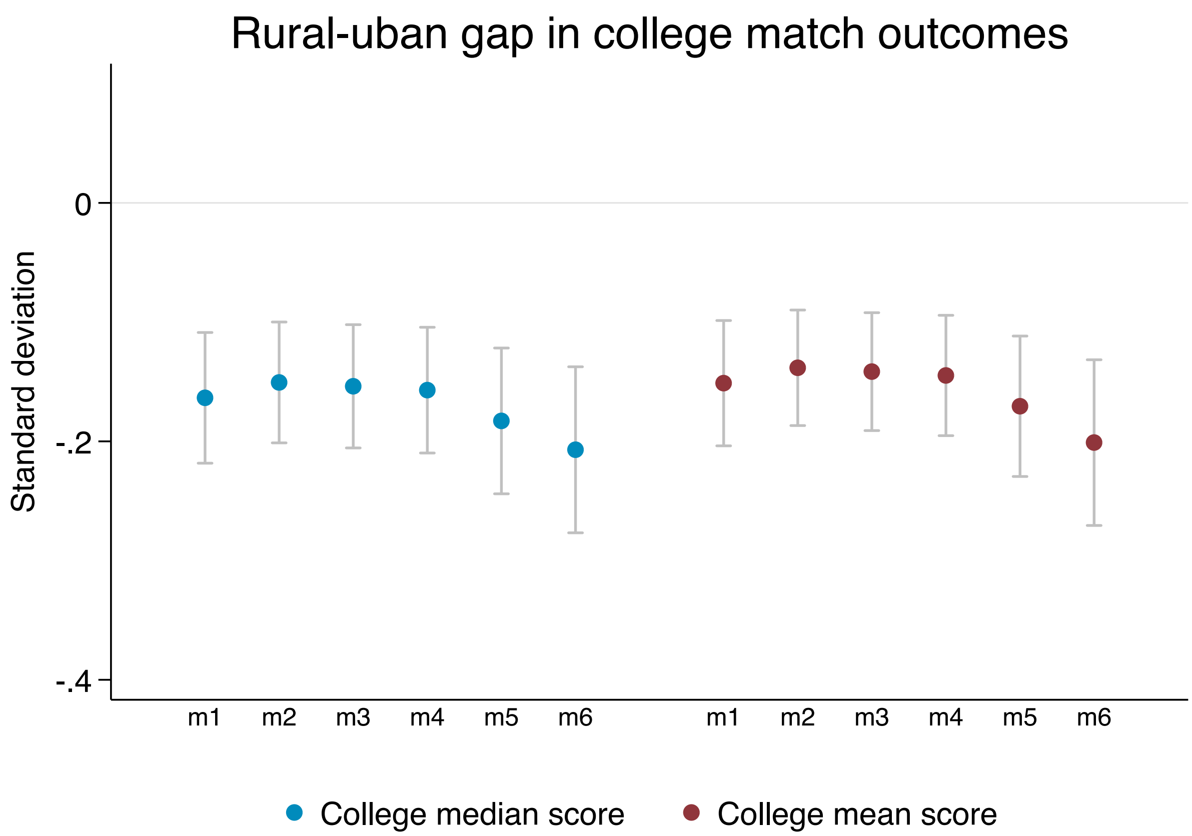


Figure 1.7. Testing the sensitivity of different college quality measures

Notes: This figure shows the consistency of the estimates using different college quality measures for students who were not admitted to any college. Results of m1 correspond to Panel D of Table 1.3 that controls for a student's CEE score and her demographic covariates. We assign the value of tier-track specific lowest college median/mean score minus 0.2 s.d. as the college median/mean score to students who are not admitted. From m2 to m5, I vary the threshold value: 0, 0.05 s.d., 0.1 s.d., and 0.5 s.d.; and in m6, I assign the lowest college median/mean score (not tier-track specific) to those students. All results (including those in the next sections, and those using other measures of the non-admitted students) remain very stable.

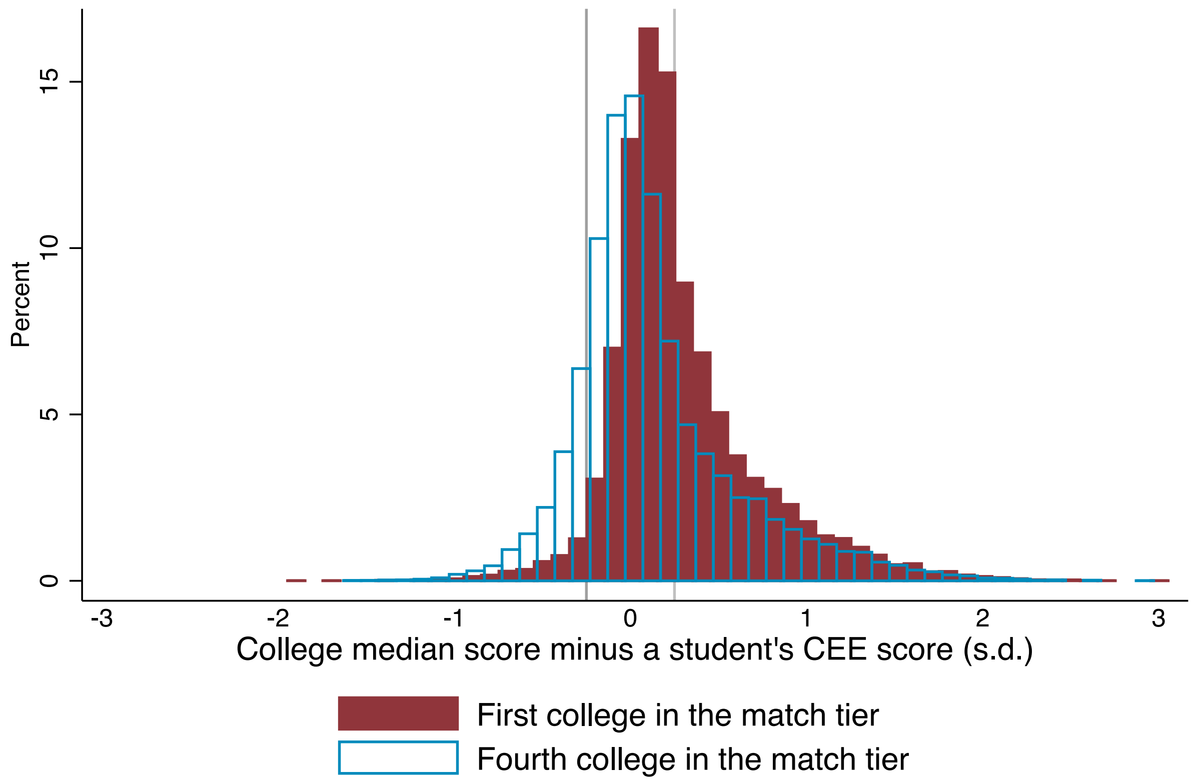
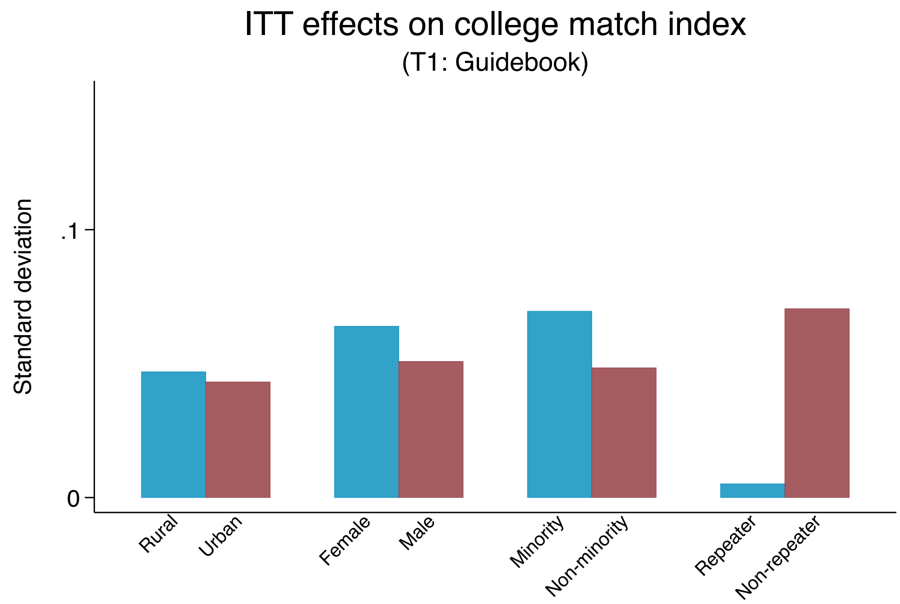
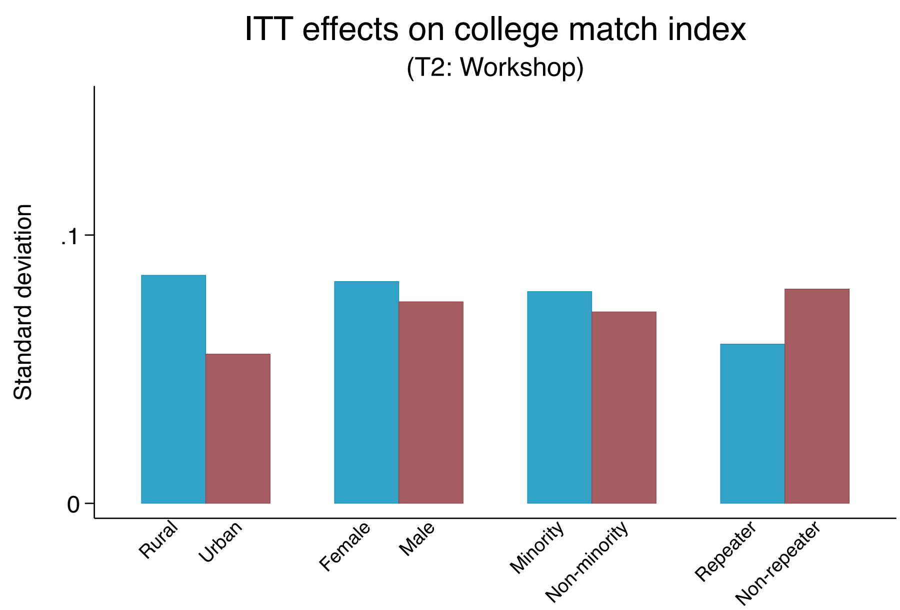


Figure 1.8. Distribution of the distance between college median score and a student's own score

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 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 primary targeting tier. Two vertical gray lines indicate the boundary of the match range (0.25 s.d. from zero).



(a) T1: Guidebook



(b) T2: Workshop

Figure 1.9. 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 1.2, but with each subsample (e.g., rural students vs. urban students) separately.

1.10 Additional Tables

Table 1.8. 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 1.9. Rural-urban gap in college access and match: Itemized outcomes of the index measure

	College median (s.d.) (1)	College mean (s.d.) (2)	Coll min (s.d.) (3)	Quality (s.d.) (4)	Ranking (pctl) (5)
Urban mean	0.017	0.026	-0.599	-0.111	53.872
	[1.192]	[1.172]	[1.362]	[1.758]	[33.252]
		A. No controls			
Rural-urban gap	-0.284*	-0.271*	-0.400**	-0.465**	-8.151*
(β_1)	(0.160)	(0.159)	(0.161)	(0.202)	(4.548)
		B. Control for CEE score			
Rural-urban gap	-0.153***	-0.141***	-0.278***	-0.314***	-4.598***
(β_1)	(0.025)	(0.024)	(0.040)	(0.055)	(0.981)
		C. Control for CEE score and demographics			
Rural-urban gap	-0.171***	-0.158***	-0.272***	-0.358***	-5.305***
(β_1)	(0.027)	(0.026)	(0.038)	(0.057)	(0.912)
		D. Control for CEE score, demographics, and class fixed effects			
Rural-urban gap	-0.099***	-0.090***	-0.144***	-0.210***	-2.986***
(β_1)	(0.016)	(0.015)	(0.023)	(0.033)	(0.480)
N	31,777	31,777	31,777	31,777	31,777

Notes: This table reports the OLS regression (Equation 1.1) results of the rural-urban gap in college match outcomes (as being summarized in the single index), using the universe of the untreated sample in 2016. The top two rows report mean and standard deviation (in square brackets) of each outcome variable for urban students. Each cell in Panels A-D is from a separate regression, showing the coefficient on the indicator of "rural *hukou* residence" (β_1). Student-level demographics include indicators of female, minority, repeater, and STEM track, and age by June 7, 2016 (=1 if at least 18 years old). **College median/mean/min scores** are constructed using all the admissions data in Ningxia in 2016. **Quality (standardized)** measures college quality using national data on college (admissions scores, inputs and employment data) from 1996-2017, and **Ranking** is the corresponding ranking percentile. Standard errors in parentheses are clustered at high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 1.10. Rural-urban gap in college access and match: Additional outcomes

	Application (=1) (1)	Enrollment in 2016 (=1) (2)	Repeating in 2017 (=1) (3)	Match (=1) (4)	Overmatch (=1) (5)
Urban mean	0.923	0.813	0.162	0.646	0.123
	[0.267]	[0.390]	[0.368]	[0.478]	[0.329]
		A. No controls			
Rural-urban gap	-0.029**	-0.100***	0.092***	-0.092***	-0.006
	(0.013)	(0.020)	(0.017)	(0.028)	(0.016)
		B. Control for CEE score			
Rural-urban gap	-0.021**	-0.087***	0.085***	-0.070***	-0.022**
	(0.009)	(0.011)	(0.013)	(0.012)	(0.010)
		C. Control for CEE score and demographics			
Rural-urban gap	-0.037***	-0.097***	0.096***	-0.075***	-0.022**
	(0.009)	(0.014)	(0.015)	(0.011)	(0.009)
		D. Control for CEE score, demographics, and class fixed effects			
Rural-urban gap	-0.021***	-0.053***	0.046***	-0.047***	-0.011***
	(0.006)	(0.008)	(0.008)	(0.009)	(0.004)
N	31,777	31,777	31,777	31,777	31,777

Notes: This table reports the OLS regression (Equation 1.1) results of the rural-urban gap in additional college access and match outcomes, using the universe of the untreated sample in 2016. The top two rows report mean and standard deviation (in square brackets) of each outcome variable for urban students. Each cell in Panels A-D is from a separate regression, showing the coefficient on the indicator of "rural *hukou* residence" (β_1). Student-level demographics include indicators of female, minority, repeater, and STEM track, and age by June 7, 2016 (=1 if at least 18 years old). **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. Standard errors in parentheses are clustered at high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 1.11. Sample description: 2016 RCT

	All (1)	Not in RCT sample (2)	RCT sample (3)	Control (4)	T1 (5)	T2 (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 1.12. Balance checks: Prediction of treatment status using student-level covariates in 2016 RCT

	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 1.13. ITT effects on additional college access and match outcomes: Guidebook and workshop in 2016

	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]
		A. Without school covariates			
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)
		B. With school covariates			
T1 (guidebook)	0.018	0.026	-0.033	0.028	0.017*
	(0.225)	(0.294)	(0.267)	(0.110)	(0.093)
T2 (workshop)	0.021	0.031	-0.051	0.009	0.018
	(0.278)	(0.309)	(0.153)	(0.676)	(0.108)
N	32,834	32,834	32,834	32,834	32,834

Notes: This table reports the OLS regression (Equation 1.2) 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. School covariates are aggregated mean values of student CEE score and demographics. 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 1.14. ITT effects on college access and match outcomes for high achieving students: Guidebook and workshop in 2016

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 1.2) 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. School covariates are aggregated mean values of student CEE score and demographics. 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 1.15. 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 1.2) 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 1.11.1. 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. School covariates are aggregated mean values of student CEE score and demographics. 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 1.16. 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 1.2) 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 1.11.1. Sample includes high-achieving 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). All regressions control for student-level covariates (CEE score and demographics) and strata fixed effects. School covariates are aggregated mean values of student CEE score and demographics. 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%.

1.11 Additional Descriptions

1.11.1 Correlate College Application Behaviors with College Admissions

Tier-Specific College Applications in Chinese Centralized Admissions

As introduced in subsection 1.2.3, 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 1.17 shows a simplified version of the college application form in Ningxia in 2016. On the one hand, the application (administrative) process is simplified. Many common requirements in decentralized admissions systems (e.g., score-sending, institution-specific essays, AP courses, reference letters) are no longer needed. Students need to choose colleges and majors of their interests from the pull-down menu in the online application system. If they already have a list of interested and majors at hand, they can finish the application process in minutes.

On the other hand, the application is complicated. Students would have to consider every cell in the application form in Table 1.17. 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).[52]

Measuring College Application Behaviors Using Actual Choice Data

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

- **[Strategy 1.1] Number of applied colleges.** <u>The behavioral rationale</u> 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 I 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. <u>I construct this variable</u> by counting the total number of all the colleges that a student

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

Table 1.17. 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 - EA	1								
	1								
Tier 1 - National Affirmative Action (Rural)	A								
	B								
	C								
Tier 1	A								
	B								
	C								
	D								
Tier 1 - Provincial AA (Rural)	A								
	B								
Tier 1 - AA (Minority)	A								
	B								
	C								
Tier 1 - Special majors	1								
Tier 2	A								
	B								
	C								
	D								
Tier 2 - AA (Minority)	A								
	B								
	C								
Tier 2 - Special majors	1								
Tier 3	A								
	B								
	C								
	D								
Tier 4	A								
	B								
	C								
	D								

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

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 I recommend students to think about their applications carefully and the number of colleges to apply to is related to the targeting strategies in the second set variables.

- **[Strategy 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 that the college-then-major admissions give each student only one college temporary admission chance. If a student is eventually rejected by a college due to the unmatched of major applications, she will not be considered by other colleges in the same institutional tier and has to move down to lower tiers. In practice, many students only have strong major preferences, but do not understand the need for this strategy to reduce their rejection risks. I construct this variable by calculating the percent of major applications over total available major numbers given the colleges that a student applied to. Sample mean is 69.9%, with a minimum of 16.7% and a maximum of 100%.
- **[Strategy 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 admissions scores than her CEE score. If that student accepts flexible major assignment within that college, then the college will assign her to a major that still has a spot (but that major may not be her interested one). The flexible assignment is actually to increase admission probability by sacrificing major preferences. I construct this variable by calculating the percent of college applications accepting flexible major assignment over the number of applied colleges. Sample mean

is 69.2% with a minimum of 0 and a maximum of 100%. The strategy, which I 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 the accurate predictions and decisions. This set of strategies are the key elements of our behavioral interventions as well as the data analysis in a students' college choice and application. Many students do not understand the underlying mechanisms of college admissions that only rank (but not raw score) matters. They naively compare their CEE score in this year with college admissions raw scores, which results in large errors of identifying college types. Students may use different strategies in different tiers, but I 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, I focus on college-level application behaviors, but those choices of majors within each college is also worth exploring in the future research.

- **[Strategy 2.1] Estimated gap (within 0.15 s.d.).** The behavioral rationale is that students should equate their CEE score to admissions 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 CEE score of 500 should then look at colleges with admissions scores around 550 in 2015. If she applied to colleges with admissions scores around 500 in 2015, she would be very much likely to undermatch. The raw scores vary dramatically over the years. Suppose that the raw CEE scores are 600 and 550 for a student ranked 10,000 in 2016 and 2015, if a student with CEE score of 600 in 2016 applied to colleges with admissions scores around 600 in 2015, she would not be likely to be admitted by an undermatched college, but being rejected by all of her applied colleges. I 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.[53] 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.[54]

- **[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. I 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.[55]

- **[Strategy 2.3] Apply to colleges without admissions data in the prior year.** The number of colleges that admit students in one province may change over time. Each year there are "new" colleges for students to apply to. The behavioral rationale is that students need to infer/predict the admissions data in previous years for these "new" colleges using other information, and they may take risks of applying to these colleges. However, if most students are risk-averse and do not apply to those col-

[53] I 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 I use other choices or a summary statistic of these choices.

[54] Figure 1.8 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.

[55] 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.

leges, it is a good opportunity for skillful students to gain an overmatched admission. I construct this variable by identifying students who applied to colleges in the match tier without admissions data in the prior year. Sample mean is 2%.

- **[Strategy 2.4] List the applied colleges in a descending order in the match tier.** The behavioral rationale is that students should apply to a mix of reach, peer and safety colleges to maximize their opportunities of getting into reach and peer colleges, and to minimize the risks of being rejected by all (Hoxby and Avery, 2013). In order to correctly identify types of reach, peer and safety colleges, students need to understand the classification of these types (a rule of thumb is a 0.05-0.15 s.d. threshold) based on score-equating. Then, for the four college choices within each tier, given the institutional feature of Differed Acceptance (Parallel) mechanism, students should list their four choices in the descending order (choice A > choice B > choice C > choice D), otherwise any choices in higher orders with higher *ex post* admissions scores are meaningless. I construct this variable by a dichotomous indicator of students who did so in their match tier. Sample mean is 31%.
- **[Strategy 2.5] Targeting.** The behavioral rationale is that, although students are nudged to apply to a mix of reach, peer and safety colleges, they should not aim too high or too low. In other words, they need to have a tight range of colleges (centering around their CEE scores). I 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 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 information and knowledge to differentiate/understand different AA programs. National AA programs are of high quality (in selective colleges), but provincial AA programs are lower-quality. I construct this variable by identifying that if a student applied to any AA programs. Sample mean is 22%, with a minimum of 0 and a maximum of 1.
- **[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. I construct this variable by identifying that if a student applied to any early admissions programs. Sample mean is 15%, with a minimum of 0 and a maximum of 1.
- **[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. I construct this variable by counting the percent of applied majors in teacher's education. Sample mean is 5.2%, with a minimum of 1 and a maximum of 40.

Student preferences and tastes are individual-specific strictly unobservable. Particularly in constrained college applications, revealed preferences may not be precisely true. I 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 college with larger admissions 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 admissions probability (Kamada and Kojima, 2015), but students may be unaware of the quota information, which is provided to them by the Department of Education. I 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 (Hillman, 2016; Hoxby, 2000; Long, 2004; Miller, 2017; Ovink *et al.*, 2018). It is also true that high-quality colleges concentrate in the economically developed regions in China. Ningxia province, as a low-income region, lacks high-quality colleges. I construct this variable by calculating the percent of applied colleges locating in out-of-province regions (excluding economically advanced regions and Ningxia's neighborhood provinces, the latter is treated as "in-province"). Sample mean is 38.8%, with a minimum of 0 and a maximum of 1.
- **[Preference 2.2] Out-of-province (in the advanced regions including Beijing, Shanghai, Guangdong) colleges.** I construct this variable by calculating the percent of applied colleges locating in the most economically advanced regions of China, including Beijing, Shanghai, Guangdong. Sample mean is 6.6%, with a minimum of 0 and a maximum of 1.

The last set of preferences are major choices. I 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.** I construct these variables by calculating the percent of each major group over the total number of applied majors. The mean values of those majors in Economics-related, Agricultural-related, Computer science-related, International-related, and Medical-related are 24.1%, 1.3%, 3.2%, 1.6%, 11.4%. I did not provide direct interventions on major choice but provided information about all the majors (e.g., coursework, college life, labor market outcomes). I 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).

Correlations Between Applications and Admissions

In Table 1.4, I extend Equation 1.1 by including the measures of our constructed strategies and preferences to examine how much they could explain the rural-urban gap in college match. Column (1) estimates the same full model controlling for CEE score and demographics as in column (2) of Table 1.3; column (2) excludes students who do not apply to college. The rural-urban gap decreases from -0.176 s.d. to -0.123 s.d., but remains statistically significant. Column (3) controls for high school fixed effects, which could capture the differential access to information and guidance between schools. The gap substantially decreases; however, it remains both economically and statistically significant.

In columns (4)-(7), I add each set of strategies and preferences (principal component factor index) stepwise. Assuming that rural students follow the same general advice with urban students, column (4) show that the gap decreases by 0.05 s.d.. Holding targeting strategies equal, the rural-urban gap largely decreases by 0.34 s.d., about half of the gap in column (4). Coefficients on the targeting strategies measures show that applying appropriately for peer, reach and safety colleges significantly improves college match. Special programs and college-major preferences do not explain much of the gap, while a few of the individual items are significantly correlated with college median score (Table 1.18), but between rural and urban group differences may be smaller than within-group individual heterogeneities. In

Table 1.18, 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. Consistently, among all the strategy and preference measures, targeting strategies explain the largest proportion of variations in college match.

Alternatively, Oaxaca-Blinder decomposition shows a very consistent story. About 0.065 s.d. of the 0.105 s.d. estimated rural-urban gap (rural: 0.082; urban: 0.187) in college median score, controlling for CEE score and demographics, is explained by rural-urban differences in the college choice strategies and preferences that I construct. The set of targeting strategies explain a 0.070 s.d. in the rural-urban gap.

I also use Oaxaca-Blinder decomposition to examine the between-group differences in applications and admissions. About 0.078 s.d. of the 0.132s.d. estimated rural-urban gap (rural: 0.058; urban: 0.189) in college match index, controlling for CEE score and demographics, is explained by rural-urban differences in the college choice strategies and preferences that I construct. The set of targeting strategies explain a 0.065 s.d. (83% of the explained differences) in the rural-urban gap. The other strategies and preferences explain very small differences (general advice: 0.008 s.d.; special program: -0.001 s.d.; location: -0.009 s.d.; tuition & quota: 0.013 s.d.; majors: 0.003 s.d.).

Table 1.18. College choices and the rural-urban gap in admissions outcomes

	Sample mean		Outcome: Index of college match			
	Rural	Urban				
	(1)	(2)	(3)	(4)	(5)	(6)
Rural-urban gap			-0.074***	-0.026***	-0.028***	-0.029***
			(0.012)	(0.007)	(0.007)	(0.007)
# of colleges applied [S1]	7.4	6.9		0.019***	0.022***	0.022***
				(0.003)	(0.004)	(0.004)
# of colleges applied$\hat{2}$ [S1]				-0.001***	-0.001***	-0.001***
				(0.000)	(0.000)	(0.000)
% of majors applied [S1]	67.3	73.1		0.000	0.000*	0.000*
				(0.000)	(0.000)	(0.000)
% flexible major assignment [S1]	63.4	76.4		0.000***	0.000***	0.000***
				(0.000)	(0.000)	(0.000)
Estimated gap within 0.15 s.d. (=1) [S2]	0.33	0.43		0.043***	0.046***	0.054***
				(0.008)	(0.008)	(0.008)
Did not apply for matched tier (=1) [S2]	0.19	0.10		-0.526***	-0.515***	-0.518***
				(0.032)	(0.033)	(0.032)
Missing prior year data (=1) [S2]	0.01	0.03		0.049***	0.049***	0.032**
				(0.012)	(0.012)	(0.012)
Descending (=1) [S2]	0.27	0.41		0.097***	0.095***	0.082***
				(0.006)	(0.005)	(0.005)
Targeting (=1) [S2]	0.32	0.45		0.035***	0.033***	0.031***
				(0.006)	(0.006)	(0.005)
Affirmative action (=1) [S3]	0.29	0.19			-0.073***	-0.048***
					(0.013)	(0.012)
Early admissions (=1) [S3]	0.16	0.17			0.046***	0.016*
					(0.010)	(0.009)
% teachers' colleges [S3]	6.3	3.8			0.002***	0.002***
					(0.000)	(0.001)
College tuition (1000 RMB) [P1]	5812	6838				-0.024***
						(0.002)
College quota [P1]	860	518				-0.000*
						(0.000)
% out of province [P2]	29.3	50.6				0.001***
						(0.000)
% advanced regions [P2]	4.5	9.1				0.002***
						(0.000)
% economics majors [P3]	22.6	26.0				0.002***
						(0.000)
% agricultural majors [P3]	1.3	1.3				0.000
						(0.001)
% CS majors [P3]	2.9	3.5				0.004***
						(0.000)
% international majors [P3]	1.2	2.1				-0.002*
						(0.001)
% medical majors [P3]	12.8	9.7				0.001***
						(0.000)
Observations			28,806	28,806	28,806	28,806
R-squared			0.665	0.724	0.726	0.733

Notes: This table reports the OLS regression (Equation 1.1) 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. [S*] denotes strategy groups, and [P*] denotes preference groups. Columns (1) and (2) report sample mean for rural and urban students. Regressions in columns (3)-(6) 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%.

1.11.2 Intervention Descriptions (Guidebook & Workshop)

The Guidebook

The "How to apply for 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, I distributed the printed guidebook to treated students through high schools (on June 20). In 2017, I 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 1.10), I 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). To increase its credibility and attractiveness, in the 2017 version, I also added the logo of Peking University.

The outline of the guidebook is as follows (Panel D of Figure 1.10):

1. Six steps in college applications
    (a) Score equating
    (b) Make use of past admissions data
    (c) Select a short list of colleges
    (d) Identify the reach, peer and safety colleges and apply to a mix set of them
    (e) Major choices within each college
    (f) Tier-specific plans (with a focus on the match tier)
2. Understanding college admissions policies
    (a) Background: Track, Tiers, Tier cutoff
    (b) Deferred Acceptance (Parallel) mechanism
    (c) College-then-major admissions
        - Major admissions 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 1.11)
        - National employment trends by majors (Panel B of Figure 1.11)

(c) Application guidelines and tips

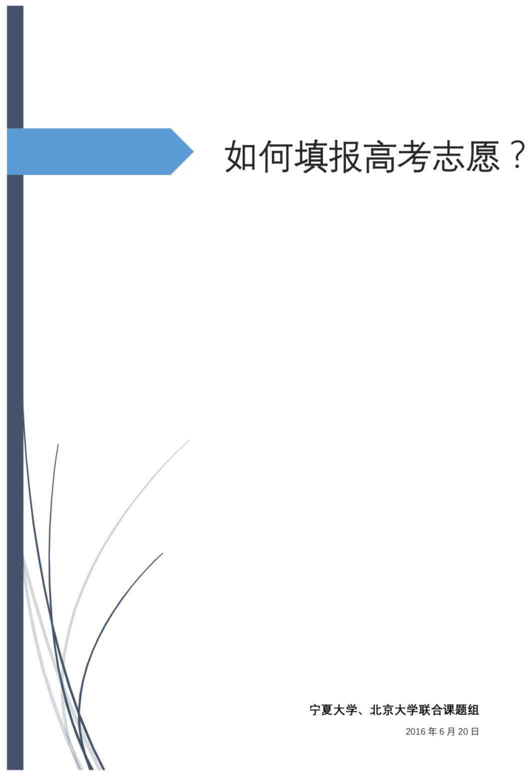
Recommended online sources. In preparing the guidebook, besides summarizing our own experience and knowledge, I have learned greatly from existing sources. Our research team carefully reviewed more than 200 Chinese websites and guidebooks that contained information about college entrance exam and college applications. I 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, I 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 1.11, I list nine "college applications" websites, each of them covers some of the information that I think is relevant to college choices and applications. From the left to the right, these information items are:

- College introduction (1)
- Schools, majors within each college (2, 3)
- College admissions guidelines (4)
- Admissions scores (5)
  - The most reliable source is the printed book provided by the provincial Department of Education; I also purchased a few copies in 2016 and 2017 for the one-on-one advising
- Housing and dinning (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)

School Workshop in 2016

I provided school workshop in seven randomly chosen high schools. Workshops were organized by local district and high school schools. To minimize the quality variations in the workshops, I 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 1.12 and Figure 1.13 show the sample pictures.



(a) Cover (2016 edition)

(b) Packages from the press, 2016, Beijing

Figure 1.10. The guidebook “How to apply for college?”

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

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 1.11. Sample contents in the guidebook “How to apply for college?”

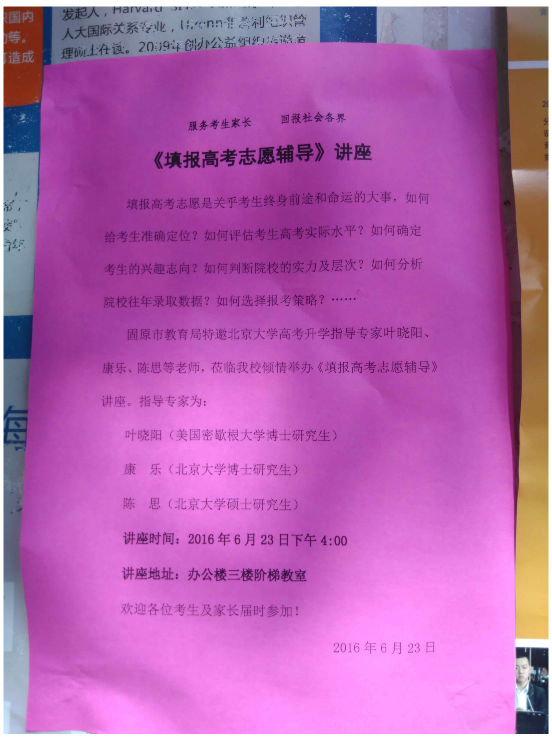
Notes: This figure shows sample contents in the guidebook. Panel A lists nine websites with a cross-tab of available information on each website that I selected from about 200 Chinese websites. Panel B shows that the employment trend graph by major that was created using data on every college graduate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

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

(b) Helan No.1 High School

Figure 1.12. High school workshops in 2016

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



(a) School poster

(b) (Late) Q&A after workshop

Figure 1.13. 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, I had a team of 3-4 members there for brief follow-up Q&A after each workshop.

1.11.3 Intervention Descriptions (One-on-One Advising)

Advising Work-Flow

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

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

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

programs, and college-major trade-offs. The decision would be based on the predictions in Step 6.

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

Examples

Figure 1.14 shows a sample picture of the one on one advising in 2016. Each shows some behavioral barriers that students had and how I helped them in the college choice process.

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

Figure 1.14. Example of the online individualized advising in 2016

Notes: This figure shows a typical case of our 1-on-1 advising. In Panel A, the student asked whether Shandong University was beyond the range of "reach college" to apply to. The advisor asked the student to do the CEE score equating and asked for the scores in the past three years (564, 588, 588). After reviewing the admissions data, the advisor replied that it was appropriate to list Shandong University as her first choice. In Panel B, the student sent a message after about one month that "Thank you for your advising. I have been admitted to Shandong University." The conversations were at QQ, one of the two largest chat forms in China.

1.11.4 Estimating the Effects of the One-on-One Advising in the 2016 Program

This appendix subsection describes how I (approximately) measure the treatment effects of the one-on-one advising in the 2016 program, which is expected to serve as a comparison benchmark for evaluating the advising programs in 2017.

Students who attended one of the 5 school workshops (randomly chosen from the total 7 T2 schools) were provided opportunities (vouchers) to receive individualized advising from experts in our research team. The one-on-one advising through the two largest online chat forms in China lasted until each student finished college applications. In addition to the general information and knowledge, the advising program was designed to help students deeply learn such college-going knowledge (like after-school coaching) and to provide customized guidance and also suggestions on preference adjustment in applications. The most crucial component of the advising program is that I helped students equate CEE scores over multiple years and apply properly for a set of reach, peer and safety colleges (*targeting strategies*). In contrast, students who received the guidebook or attended the workshop only learned the general knowledge and skills, but they needed to do their own work of using the knowledge in college choice and application. Table 1.19 shows that the take-up of the advising program is about 1.5 percent in the whole sample (N=119), and about 3 percent of high achieving students who received the voucher eventually participated in our advising program (N=72).[56]

The primary purpose of the advising program in 2016 is to examine whether our individualized advising actually works and to understand how it works through individual cases. It is seen as a supplemental service for students who attend the workshop. This design makes it nearly impossible to formally estimate the treatment effects of the advising program because

[56]More than 800 hundred users contacted us, but the screening process (based on school and student IDs) largely decreased the actual take-up. I only provided general advice (same as in the guidebook or workshop) to those unverified students. As shown in Table 1.19, there is a small and insignificant spill-over effect on take-up among students in other schools.

students who were offered the voucher but did not participate (some of the never-takers) may have received the workshop treatment. Therefore, I present some exploratory results to gain the sense of its effectiveness, which are not strictly causal.

I first examine the observed admissions results of our "treated" students. Consistent with the descriptive results shown in Figure 1.4, Panel A of Table 1.20 shows similar results by differentiating the naïve effects between students who received the 1-on-1 advising program and those who were assigned to the guidebook-workshop combined intervention (T2). Results show that on average, students who received the advising had better college access and match outcomes, holding CEE score and demographics equal. Note that the estimated naïve effects are largely biased downward.

I further take advantage of the fact that I randomly provided the advising vouchers to five of the seven T2 schools to approximately quantify the individualized advising program effects. I start with a thought experiment. Under the strong assumption of treatment effect homogeneity, I assume that the effect of guidebook-workshop is the same across schools. The estimated ITT effects of the guidebook-workshop intervention in the five voucher schools include the effects from both the workshop and the advising program, but the ITT effects in the two no-voucher schools are only from the workshop intervention. Then I could derive the treatment effect of the advising program using the Wald estimator:

$$\widetilde{TOT}_{advising} = \frac{ITT_{workshop\&voucher} - ITT_{workshop\&novoucher} * (1 - Prob_{advising})}{Prob_{advising}} \tag{1.3}$$

where $Prob_{advising}$ is the take-up rate.

In panels B and C of Table 1.20, using the experimental design, I separately estimate the ITT effects of $ITT_{workshop\&voucher}$ and $ITT_{workshop\&novoucher}$ for each outcome among high achieving students. Using the take-up rate of 3% from Table 1.19, I calculate the approximate treatment effects of 1-on-1 advising. The results are consistent using different outcome measures that the advising program increases admissions by 13.1 percentage points, admitted college peer median score by 0.706 s.d., admitted college quality by 0.974 s.d. and 11 rank percentiles, and the college access and match index by 0.210 s.d..[57] Overall, the results provide suggestive evidence that our "human instruction" individualized advising largely improved treated students' college access and match, particularly among high achieving students. I should note again that the approximate estimate builds on strong assumptions. It is not the rigorously causal effect of the advising program, but it presents a benchmark reference.

[57] The effects for lower-achieving students are smaller (and even negative). This is mainly because a few lower-achieving students finally chose to retake the CEE in the next year in order to go to higher quality colleges.

Table 1.19. Take-up of the individualized advising program

	Received 1-on-1 advising			
	All students		High achieving	
	(1)	(2)	(3)	(4)
T2 (workshop & voucher)	0.015***	0.015***	0.032***	0.030***
	(0.002)	(0.003)	(0.004)	(0.006)
T2 (workshop & no-voucher)	0.003**	0.003	0.006	0.006
	(0.001)	(0.002)	(0.004)	(0.004)
T1 (guidebook)	0.001	0.002	0.004	0.004
	(0.001)	(0.001)	(0.004)	(0.004)
Rural		-0.001		-0.001
		(0.001)		(0.003)
Female		0.001		0.003
		(0.001)		(0.003)
Minority		0.001		0.002
		(0.001)		(0.002)
Age		0.000		-0.001
		(0.000)		(0.001)
STEM		0.002*		0.003
		(0.001)		(0.003)
Repeater		-0.002		-0.002
		(0.002)		(0.005)
CEE score		0.004***		0.009
		(0.001)		(0.005)
Observations	32,834	32,834	7,977	7,977
R-squared	0.009	0.014	0.018	0.020

Notes: This table reports the OLS regression results of the take-up of individualized advising program in 2016. All regressions control for strata fixed effects. Standard errors in parentheses are clustered at high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 1.20. Effects of the individualized advising program

	Sample: High achieving students				
	Admissions	Index	College median	College quality	College ranking
	(1)	(2)	(3)	(4)	(5)
		A. Naïve effects			
T3 (1-on-1)	0.003**	0.093**	0.074*	0.086**	1.512**
	(0.001)	(0.036)	(0.038)	(0.035)	(0.559)
T2 (workshop, no T3)	0.001	0.077***	0.053***	0.063***	1.187***
	(0.002)	(0.021)	(0.019)	(0.021)	(0.331)
T1 (guidebook)	0.001	0.057***	0.040**	0.049**	0.786**
	(0.001)	(0.020)	(0.019)	(0.022)	(0.296)
		B. ITT effects (excluding workshop & voucher schools)			
T2 (workshop)	-0.002	0.077***	0.039*	0.041	0.993**
	(0.002)	(0.021)	(0.021)	(0.029)	(0.383)
T1 (guidebook)	0.001	0.059***	0.040**	0.046**	0.741**
	(0.001)	(0.019)	(0.019)	(0.022)	(0.301)
		C. ITT effects (excluding workshop & no-voucher schools)			
T2 (workshop)	0.002	0.081***	0.059***	0.069***	1.286***
	(0.002)	(0.022)	(0.017)	(0.019)	(0.329)
T1 (guidebook)	0.001	0.057***	0.040**	0.049**	0.795**
	(0.001)	(0.020)	(0.019)	(0.021)	(0.289)
Estimated TOT effect of 1-on-1	***0.131***	***0.210***	***0.706***	***0.974***	***10.760***

Notes: This table explores the effects of the advising program. See the text for more information. All regressions control for student covariates and strata fixed effects. Standard errors in parentheses are clustered at high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

CHAPTER II

Human Instruction vs. Machine Learning: Experimental Evidence on Improving College Access and Match at Scale

2.1 Introduction

To improve college access and match, the past decade has witnessed emerging research and policy efforts to utilize behavioral interventions - low-touch information provision and intensive personalized advising programs - to help students make their optimal decisions.[1] This is because low-income students often make suboptimal college choice during the complex transition from high school to college, which results in being admitted to an academically undermatched college (Bowen *et al.*, 2009; Hoxby and Avery, 2013; Smith *et al.*, 2013; Dillon and Smith, 2017b). Undermatched college choice offsets many policy efforts to increase K-12 school quality and financial aid for these low-income and other types of disadvantaged students (see recent summaries in Glewwe and Muralidharan, 2016; Page and Scott-Clayton, 2016; Fryer, 2017). More importantly, college undermatch has large, negative impacts on students' college and labor market outcomes.

Two key policy questions remain unanswered despite the rapidly growing literature on

[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 J-PAL (2018).

using behavioral policy interventions to improve college access and match. First, as most studies focus on the decentralized admissions systems in the United States and Canada, we know little about what works in other contexts (Dinkelman and Martínez, 2014; Hastings *et al.*, 2018; Peter *et al.*, 2018). The prevalence of centralized college admissions with simplified application processes,[2] and additional institutional barriers in many developing countries, call into question the effectiveness of interventions that have shown promise in the current literature. Second, we know that personalized advising and assistance may be more effective than low-touch information provision, but we do not know much about successful scale-up approaches for the high-cost and often labor-intensive advising programs in both the decentralized and centralized systems.[3]

I addressed the first question in Chapter I. I find that low-income students in a centralized college admissions system significantly undermatch. While centralized admissions have simplified application processes and centrally-provided information, low-income students still lack sufficient knowledge and skills to make informed college choices. A behaviorally-designed college choice guide improves students' college access and match by shaping their college choice behaviors. The intensity/touch of intervention increases its effectiveness. One-on-one personalized advising that targets a student's behavioral barriers is much more effective than providing a print guidebook or a school workshop.

This chapter addresses the second question: *How to scale up the personalized college application advising program to improve student-college match and reduce inequality in college*

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

[3]Low-touch and low-cost informational interventions can be provided to a large number of students (see Hoxby and Turner, 2013; Bird *et al.*, 2017; Bergman *et al.*, 2019; Hyman, 2019), but may not be sufficiently effective. In contrast, personalized advising may be very effective, but may not be scalable (Bettinger *et al.*, 2012; Carrell and Sacerdote, 2017; Oreopoulos *et al.*, 2017; Oreopoulos and Petronijevic, 2018; Bettinger and Evans, 2019).

access? The scale-up problem is common for many social policy programs, and results from the same underlying constraint: Some "inputs" to a program are in limited supply (Davis *et al.*, 2017; Muralidharan and Niehaus, 2017). The key input to college-going advising programs - advisers - are inelastically supplied in both quantity and quality, and often come with high costs. Successful college-going intervention requires intensive guidance, instruction, and assistance (human instruction), particularly in centralized systems where strategy and sophistication are extremely important. Personalized advising is unlikely to be scalable as a system-level policy intervention to a large number of students because we may not be able to hire and train enough experienced advisers, counselors, or near-peers to provide such advising.

I propose two policy solutions to scale up individualized advising programs through the introduction of machine learning and the expansion of conventional human instruction. Both approaches have the potential to increase school effectiveness in improving college access and match without overly burdening current school resources. The machine learning approach increases the intensive margin of labor supply by using big data and machine learning algorithms to simplify the prediction of admissions probability - the most complex and time-consuming process in personalized advising. The human instruction approach increases the extensive margin of labor supply by developing a new pay-for-performance policy to incentivize high school teachers to act as temporary front-line counselors. Teachers were provided an incentive contract stipulating performance pay - about 1-2 months' salary for eligible teachers - based on class-level performance in college admissions, holding college entrance exam score constant.

To estimate the causal effects of the two proposed scale-up solutions on college access and match, I conducted a set of large-scale randomized controlled trials (RCTs) among the universe of public high school graduates in 2017 in Ningxia, one of the poorest provinces in China. I continued the collaboration with the provincial government and high schools in the *Bright Future of China Project-Ningxia*, as described in Chapter I.[4] China is an ideal setting to study behavioral interventions in college choice. First, college choice and admission in China combines the institutional features of both centralized systems and developing countries. China has the largest college admissions market in the world. Every year around ten million Chinese high school graduates apply to colleges. According to the estimates in Chapter I, more than three million students undermatch because of their college choices. Furthermore, the centralized system allows us to credibly identify the impacts of college choice behaviors on admissions outcomes, because the variation in admissions outcomes is solely determined by students' college choice behaviors when holding their college entrance exam scores equal.[5]

Both student-level and teacher-level stratified randomization designs were used to test the effectiveness of the two proposed solutions. Table 2.1 summarizes the experimental design. I randomly assigned all the students into one of the three groups: (1) 5,647 students were provided access to the machine learning assisted personalized advising, which we refer to henceforth as the machine learning group; (2) 5,370 students were provided access to a low-touch "business as usual" advising group; and (3) the remaining 43,038 students served as the control group.[6] In the machine learning group, expert advisers provided students with the conventional personalized advising with the assistance of machine learning predictions and relevant data analytics. The "business as usual" advising provided brief college application guidelines. Independently, I randomly assigned a group of 184 classroom head teachers to either the treatment group that consisted of 88 teachers, who each received a pay-for-performance contract, or to the control group, which consisted of 96 teachers. The incentive contract based on class-level college admissions outcomes was supplemented with online

[4]In 2017, we discontinued the guidebook and workshop interventions because the all the research team members were preparing for the individualized advising program. The methodological reason of this decision was that school-level randomization limits the statistical power of impact evaluation.

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

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

training sessions and other support materials.

This chapter reports four sets of results. First, using student-level randomization, results indicated that the personalized advising intervention assisted by machine learning improved college access and match, which closely mirrored the expert human instruction intervention effect in 2016; however, machine learning predictions largely replaced human labor in performing the data analysis tasks. Holding College Entrance Exam (CEE) scores and demographics equal, machine learning advising increased the probability of college admissions by 24.4 percentage points (treatment-on-the-treated; $p<0.1$). It also increased the quality of a student's admitted college by 0.598 standard deviations ($p<0.05$). The estimated effects were robust to using a family of college access and match measures. As a placebo test, a "business as usual" intervention mimicking a typical for-profit (and perhaps non-expert) consulting service did not impact admissions outcomes. One possible reason is that the low-touch intervention reduced students' intrinsic motivation and did not provide an effective guide for choosing match colleges. Results demonstrated that the treated students in the "business as usual" intervention group were less likely to use the targeting strategies. Those students were also less likely to enroll at college on time.

Next, I analyzed college choice behaviors using the unique data of students' full college applications. I show compelling evidence that the machine learning advising nudged the treated students to use the appropriate data-based college application strategies, including the targeting strategies and the general guidelines. Treated students in the machine learning group were statistically significantly and substantially more likely to make strategic college application decisions, as measured by a set of variables characterizing college choice behaviors. This improvement was consistent with the fact that treated students were more likely to apply for colleges in the recommendation list based on the machine learning predictions. Importantly, neither the machine learning intervention nor the business as usual intervention altered student preferences such as tuition, location, and choice of majors. These results suggest that a behavioral intervention in college choice may work through improving application strategies without impacting preferences for colleges and majors.

This chapter also presents novel evidence on the positive correlation between the admissions outcomes and the amount of time spent making college choice decisions. Using newly available data on the date and time of students' college application submissions, I found substantial heterogeneity in the time devoted to college choice by income, race, and gender. Students who submitted applications later had on average better college access and match outcomes. This occurred primarily through their improved college choice behaviors, particularly in the use of data analytics for predictions. The machine learning intervention increased the application time for treated students by about 27 hours (treatment-on-the-treated; $p<0.05$) during a 98-hour college application period. The "business as usual" advising decreased students' time use.

In the final analysis, I tested the effectiveness of teacher incentives to increase the supply of counselors. The limited supply of school counselors is a primary reason why schools do not effectively help students make their best college choices (Perna *et al.*, 2008, Page *et al.*, 2014). I developed a new pay-for-performance policy, which was designed to incentivize high school teachers to act as temporary front-line counselors. However, I found a statistically zero impact of advising performance incentives for teachers on students' college admissions outcomes. Descriptive evidence from survey data suggested that the reason for the null effect was likely due to teachers' lack of college choice knowledge to make informed college choices. Moreover, both teachers and students reported that teachers had limited impact on a student's college choice.

This chapter makes several contributions to the literature and to policy. First, this chapter contributes toward filling the gap of limited evidence on college-going interventions from centralized admissions systems or developing countries. Existing literature concentrates on the higher education markets in the U.S. and Canada; see summaries in White House (2014); Page and Scott-Clayton (2016); French and Oreopoulos (2017); J-PAL (2018). This chapter provides new evidence on the impact of college application assistance from the largest centralized college admissions market in the world. Centralized admission is widespread across countries in both K-12 and higher education. While it streamlines and simplifies the application process, it may require strategies and sophistication in decision-making.

The interventions in the *Bright Future of China Project-Ningxia* build on many prominent approaches, including information provision (Hoxby and Turner, 2013; Goodman, 2016; Peter and Zambre, 2017; Herber, 2018), text message reminders (Castleman and Page, 2015), and most importantly, personalized advising/counseling (Bettinger *et al.*, 2012; **?**; Carruthers and Fox, 2016; Carrell and Sacerdote, 2017; Oreopoulos *et al.*, 2017; Page *et al.*, 2019; Castleman and Goodman, 2018). This chapter shows compelling evidence that a behaviorally-designed, intensive intervention using a combination of customized information and personalized assistance substantially improves students' college choice behaviors (particularly, their use of strategies) and thus their college access and match outcomes. Moreover, I also show that taking sufficient time to form a thoughtful college choice plan improves students' college access and match. This suggests that nudging students to "think slow" may have desirable behavioral consequences (Kahneman, 2011; Heller *et al.*, 2017).

Second, this chapter links to the broader literature on the scale-up of social programs. The scale-up problem is pervasive in nearly all social programs. Personalized advising programs are effective in improving college access and match, but they are not easily scalable. Existing intensive college counseling studies have only covered a small number of students. For example, Carrell and Sacerdote (2017) provide a college coaching/mentoring program at the cost of $300 per student, but the program has only 871 treated students in six high school graduation cohorts. Oreopoulos *et al.* (2017) evaluate the Pathways to Education Program in Canada that has served only 1,274 students in 8 years. This chapter proposes and tests both the intensive margin and the extensive margin solutions to increase the labor supply of college choice "human instruction" services, which may be applied in addressing many other education policy problems.

Third, this chapter contributes to the recent literature on applying machine learning to prediction policy problems (Kleinberg *et al.*, 2015; Mullainathan and Spiess, 2017). When past data are available to learn from, the link between predictions and decisions is clear, and expertise in understanding the decision process is applied, machine learning algorithms show considerable potential for improving predictions and productivity (Chalfin *et al.*, 2016; Kleinberg *et al.*, 2017).[7] Data-based decision-making has been rapidly growing in both K-12 and higher education, such as computer/technology-assisted instruction (Muralidharan *et al.*, 2018; Taylor, 2018), digital tutoring (Burch *et al.*, 2016), learning analytics (Daniel, 2015), and predicting college application, enrollment, and success (González and DesJardins, 2002; Herzog, 2006; Acharya and Sinha, 2014; Aulck *et al.*, 2016). Oreopoulos and Petronijevic (2018) study technology-based college coaching using online exercises and text and email messaging. They find no effects of the technology-based intervention and conclude that "future technology-based interventions should aim to provide proactive, personalized, and regular support." Machine learning or data-based prediction actually has the potential to offer personalized assistance, perhaps better than human experts. Lechner and Smith (2007) examine the efficacy of caseworkers in allocating individuals to government programs, and find that statistical treatment rules do substantially better. Burkhardt *et al.* (2018) find that medical school enrollment predictions using the enrollment management model were at least as accurate as the expert human estimates, and in specific populations of interest more accurate. This chapter adds experimental evidence on the effectiveness of machine learning to improve prediction and advising in the college-going behavioral interventions.

Lastly, I provide new evidence to the literature on school managerial policies. Evaluating and rewarding teacher performance is one of the fundamental personnel policies in schools that affect school effectiveness and organizational quality (Neal, 2011; Finan *et al.*, 2017). There has been growing interests among global policymakers and school leaders in

[7]Machine learning is not a master key to all policy prediction problems without human expertise in understanding the problems. For example, McKenzie and Sansone (2017) study the prediction of outcomes for entrants in a business plan competition in Nigeria and find that machine learning methods do not offer noticeable improvements. The main reason is that the overall predictive power of both human judges and prediction models is very low, which means the key variables of the decision-making are not being measured.

providing financial incentives to schools and teachers on the basis of student learning outcomes. However, existing evidence, even well-identified using randomized experiments, has been limited and is also mixed. A few studies report positive impacts of different performance pay plans (e.g., group or individual based, pay for gains or percentiles) on student test scores in India, the U.S., and China (Muralidharan and Sundararaman, 2011; Dee and Wyckoff, 2015; Imberman and Lovenheim, 2015; and Loyalka *et al.*, 2019). However, Springer *et al.* (2012), Fryer (2013), Goodman and Turner (2013), and Barrera-Osorio and Raju (2017) find no statistically significant effects on student performance or teacher behaviors in the US and Pakistan.[8] The null effect may be attributable to the incentive structure designs. Additionally, performance pay policy (broadly, the evaluation and accountability pressures) may motivate teachers to expend effort toward short-term reward-formula outcomes such as "teaching to test" (Glewwe *et al.*, 2010) and other unintended, strategic consequences such as cheating (Jacob and Levitt, 2003), retaining students (jac, 2199), or focusing exclusively on marginal students (Neal and Schanzenbach, 2010). The null result in this chapter suggests that incentive policies alone may not be effective enough to change the input and behavior of teachers. Future work could explore the complementarities between incentives and other school policies such as accountability (Deming and Figlio, 2016; Mbiti, 2016), organizational structures (Roderick *et al.*, 2011), school finance inputs (Mbiti *et al.*, 2018), and school leadership (Springer *et al.*, 2012).

The rest of this chapter is organized as follows: Section 2 introduces the background information. Section 3 presents a brief theoretical framework. Section 4 describes the interventions. Section 5 describes our data, experimental design, and econometrics. Section 6 reports the results. Section 7 offers discussions and policy implications. Section 8 concludes.

[8] De Ree *et al.* (2018) find that a large pay increase in Indonesia improved teachers' satisfaction with their income but led to no improvement in student learning outcomes.

2.2 Background

This section describes the institutional background of centralized college choice and admission in general, and the Chinese policy context in particular. Chapter I provides detailed discussions about the different roles of information and knowledge in the college choice model, and the reasons why centralized college admissions require more strategic and sophisticated applications than decentralized admissions. Perna (2006) conceptualizes that four contextual layers shape a student's college choice decisions: the individual's habitus (e.g., demographics, social and cultural capital); school and community context; the higher education context; and/or the social, economic, and policy context. This chapter studies the poverty gap in college choice as a result of the difference between poor and non-poor students in individual's habitus (e.g., information and assistance) and school context (e.g., available resources and structural support), and how we could mitigate the poverty gap.

2.2.1 Why Do Some Students Make Suboptimal College Choices?

College choice, whether and where to go to college, is one of the most high-stakes decisions in life. Every year, millions of high school graduates all around the world make their college choices through either a decentralized system like those in the U.S. and Japan or a centralized system with national college entrance exams like the ones in Chile, China, and the UK. For the former, the application process is highly complex including SAT or ACT tests, high school transcripts, recommendation letters, personal statements, and interviews. In contrast, the centralized system largely simplifies the application process. After taking the national entrance exam, students only need to submit a rank-order application list of colleges.[9]

This chapter focuses on a student's decisions about which colleges to apply to, including

[9] Consistently, mandatory entrance exam and automatic score sending in decentralized systems improve college access and enrollment (Klasik, 2013; Bulman, 2015; Hurwitz *et al.*, 2015; Pallais, 2015; Goodman, 2016; Hyman, 2017; Hurwitz *et al.*, 2017).

the search and choice phases in the three-phase model as proposed by Hossler and Gallagher (1987).[10] The college choice model, based on the human capital theory, assumes that students are rational and forward-looking in that they maximizes their expected utility in making their choices from a feasible set of colleges (e.g., Manski and Wise, 1983; Kane, 1999; Long, 2004; Perna, 2006; Jacob *et al.*, 2018). According to this standard model, students' college choice decisions are affected by academic ability, financial resources (e.g., family income, tuition, financial aid), and individual preferences for college (and major) attributes (e.g., selectivity, college type, cost, distance, and consumption amenities).[11]

The human capital theory assumes that all the college choice decisions are *ex ante* optimal, and students make their rational and best choices with available information, even when the information is incomplete or inaccurate (DesJardins and Toutkoushian, 2005). Having better information to evaluate the benefits and costs of college attendance ensures that students make better decisions. Smith *et al.* (2013) and Dillon and Smith (2017b) suggest that students who have better access to information on college options and the college going process (e.g., from parents, networks, and schools) are less likely to undermatch.

In contrast, the behavioral theories imply that students may not always make utility-maximizing decisions (Thaler and Sunstein, 2008; Castleman *et al.*, 2015b). In particular, research in developmental psychology and neuroscience suggests that the executive brain function of children and adolescents has not fully matured. They are more susceptible than adults to behavioral barriers such as lack of self-control, automatic or short-sighted decisions, and limited attention when evaluating decisions with long-term consequences (Lavecchia *et al.*, 2016). Hoxby and Avery (2013) find that, while high-income students generally follow expert's advice to apply to a mix of reach, peer, and safety colleges, the vast majority of

[10] College choice includes several stages. For example, Hossler and Gallagher (1987) propose a three-phase model consisting of predisposition, search, and choice; DesJardins *et al.* (2006) jointly model the application, admission, financial aid determination, and enrollment decision process.

[11] 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, single-sex or coed, religious affiliation), location, campus and housing, majors and learning environment, sports and activities, academic credit, paying, additional support programs, and diversity.

low-income high achievers do not apply to any selective colleges.

One key policy issue in the transition from K-12 school to college is the persisting inequality in college access and match. Disadvantaged students (e.g., low-income, under-represented minority, first-generation) have less access to information and assistance such as school-based support and counseling that can aid in optimal choices. Conventional K-12 and higher education policies emphasize the adequacy of academic preparation and financial aid for disadvantaged students, but they are not sufficient. The emerging evidence from decentralized college admissions in North America suggests that relatively inexpensive information provision, and application process simplification, can substantially improve college access and match. These two policy levers have already been institutionalized in the centralized systems.[12] However, disadvantaged students in centralized systems still face multiple behavioral barriers in making their optimal decisions, which calls for new intervention designs.

2.2.2 Why Do Centralized Admissions Need (More) Strategic Applications?

Students need both information and knowledge about college choice strategies in order to construct a thoughtful application list with a set of reach, peer/match, and safety colleges. This process should impose both the costs of searching relevant information for college choice decisions, and the costs of learning decision-making strategies to use the information. The centralized college admission system largely reduces the search cost through centralized information provision and application process simplification. However, the decision on which colleges to apply to is still complicated. The need for highly accurate predictions of *ex ante* admission probabilities for each college in centralized systems results in substantially higher

[12]The centralized admission mechanism, considered to improve efficiency, welfare and match (Gale and Shapley, 1962, Balinski and Sönmez, 1999; Abdulkadiroğlu and Sönmez, 2003), has long been adopted in many markets, including college admissions in many countries and in some U.S. K-12 school choices (Abdulkadiroğlu *et al.*, 2005; Pathak and Sönmez, 2013; Machado and Szerman, 2017)

learning costs compared with decentralized systems.

The student-college centralized matching market consists of a set of students and a set of colleges with admissions quotas. Students submit a rank-order list of colleges. Colleges rank students using their entrance exam scores. The admission result is solely based on a student's rank-order list and entrance exam score. Students select and rank a set of colleges to maximize their expected utility. Students derive zero utility if a college rejects them. A wise college choice strategy is a function of students' test scores, their preferences and valuations for each college based on complete or incomplete information, their predicted admission probability, and other individual idiosyncratic factors. In particular, students have to make an accurate prediction of the admission probability for every considered college, which requires an understanding of the admission policy, decision-making skills and strategies, and the correct use of admissions data in past years.[13]

Several common institutional features and barriers in centralized admissions systems emphasize the role of accurate prediction and strategic application. First, the admissions are often implemented using an application list with a restricted length in many real-world examples (Arslan, 2018; Chen and Kesten, 2017). For example, students in Ningxia (the study sample in this chapter) could only apply to four selective colleges and six majors within each college.[14] Under the Deferred Acceptance mechanism, but with restrictions on the application, truthful revelation of preferences is no longer a good strategy. Second, many centralized matching systems only match subjects (e.g., students) to *at most* one option, which imposes a high risk of all applications being rejected when students aim too high. Students have to evaluate the admission probability for each college based on admissions outcomes (cutoff or median scores) in prior years. Third, students must rank their

[13]This entire set of the requirements are summarized as "the knowledge of college choice" in Chapter I. Knowing the experts' advice about applying to a set of reach, peer, and safety colleges (as described in Hoxby and Avery, 2013) is not enough, because identifying which colleges are reach, peer, or safety requires a student's understanding of all the "knowledge" and intensive data analysis as well.

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

applications in order. Without accurate information about college quality (or other individual preferences), and without precise predictions of admission probabilities, students may make mistakes in ordering (e.g., listing a safety college at the first choice, or listing a reach college at the last choice). Lastly, centralized admissions may operate in a very short period of time. In order to do a thorough search and assess college fit, students have to search for and analyze a large volume of college/major information (thousands of colleges) from many reliable or even unreliable sources.

Chinese college admissions. China has a province-level centralized college admissions system. In early June, students take the annual College Entrance Exam (CEE). Students then compete with peer applicants within the same province and STEM/non-STEM track for college-major spots that are predetermined by each college across the country.[15] Students submit their college application lists (typically 4-10 colleges in each institutional tier, 4-6 majors for each college) in the Department of Education online system.[16] Admission is solely determined by students' CEE scores and their applications.

College application and admission proceeds by institutional tiers.[17] Students can only apply to colleges for tiers in which their CEE scores are above the tier cutoffs. There are also special programs such as income and racial affirmative action, within each tier. A student can apply to more than 50 colleges (she does not have to) or to none. Most students who do not apply to any college choose to retake the CEE in the next year. Each student is matched with *at most* one college-major through a parallel mechanism (like the Deferred Acceptance mechanism, as discussed in Chen and Kesten, 2017). Students who decline the admission offer or are not admitted to any college either retake the CEE after one year, or go to the

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

[16]All of the relevant information, including college admissions results in prior years, and information about tuition, location, and quota are publicly provided to students by the Department of Education. We show and explain a typical Chinese college application form in subsection 1.11.1.

[17]Tier 1 includes the nation's elite colleges. Only high-achieving students are eligible to apply. Tier 2 and 3 consists of non-elite public and private four-year colleges, respectively. Tier 4 includes three-year vocational colleges, which resemble community colleges in the U.S.

labor market.

The basic application strategy is similar for Chinese students as for American students. They have to apply for a set of reach, peer, and safety colleges to maximize their opportunities of being admitted by a reach or peer college, and to minimize their chances of being rejected by all of the colleges to which they have applied. Since college admission is uncertain and risky - with a limited number of choices and the fact that each student will be admitted to at most one college - students must "game" the college application strategically with accurate predictions. Additionally, students have to choose college and major simultaneously. The match process is college-then-major, which complicates the college choice decision-making process, even though the application process itself is simple and most information is available to every student.

2.2.3 Why Did We Create the Bright Future of China Project?

Together with collaborators, we started the *Bright Future of China Project* in 2015 to study policy solutions for improving college choices and admission outcomes for low-income and other types of disadvantaged students in the Chinese centralized admissions system.[18] My earlier work motivated the development of the project. A large proportion of low-income students in the Chinese centralized system, like those in decentralized systems, make sub-optimal college choices (Loyalka *et al.*, 2017), even though the centralized college application process is not as complex as the one in decentralized systems. Meanwhile, I have been learning much from the emerging research and policy efforts to support low-income students in the transition from high school to college in the decentralized systems, mostly in the United States and Canada. Working with a research team hosted by the China Center for Education and Human Resources Research, we aim to use large-scale randomized experiments to explore effective, behaviorally-informed interventions to improve low-income

[18]We 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.

students' college access, choice and success at scale.[19]

To explore effective school-based interventions for low-income students, I have collaborated with local governments and high schools in several of the most underdeveloped provinces in China. We conducted the first-year program in Ningxia province in 2016, which serves as a benchmark proof of intervention effectiveness for the follow-up programs. Chapter I provides clear evidence that the knowledge-based interventions substantially improved college choices and admissions for low-income students. This chapter studies the 2017 follow-up experimental designs, which test the scale-up solutions for expanding the labor-intensive personalized advising intervention.[20] Table 2.1 provides a summary of the experimental designs in 2017.

Ningxia is one of the poorest provinces in China. Figure 2.8 shows the geographic location of Ningxia - in northwestern China. In 2017, the annual per capita disposable (after tax) income of urban residents was about \$4,200 (national average: \$5,600), and the figure for rural residents was \$1,650 (national average: \$2,060). About 800,000 of Ningxia's 6 million population are under the extreme poverty line, earning less than \$1 a day. Moreover, more than 35% of its population are Muslim minorities. Ningxia is a very typical low-income province regarding college applications and admissions. Each year, nearly all high school graduates (about 60,000) take the College Entrance Exam.[21] Around 90% of the exam takers apply to college, 85% are admitted to college, and fewer than 10% are admitted to elite colleges.

[19]Both of the two co-PIs of this project have decade-long intensive experience of advising college choices. As introduced in Chapter I, we brought together education policy researchers, experienced high school counselors, and college admissions officers to develop the college choice and application guide and to provide personalized advising and assistance.

[20]Figure 2.7 provides the project timeline of the *Bright Future of China-Ningxia* Project in these two years. We have not yet received access to the administrative data in the other two provinces (Anhui, Yunnan) in which we conducted similar interventions in 2017.

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

2.3 Conceptual Framework

This section outlines the mechanisms through which the designed advising program could improve college choices and admission outcomes, and the potential impacts of different scale-up designs. College choice is complicated in three layers: information, the application process, and strategy. The centralized college admissions system simplifies the behavioral barriers for information and in the application process. Accordingly, the human instruction advising program in the *Bright Future of China Project* targets students' barriers in the use of strategic and sophisticated college applications.[22] Following Lavecchia *et al.* (2016), I target three groups of behavioral barriers: (1) some students make mistakes with little information or with many options, (2) some students rely too much on routine, and (3) some students focus too much on the present.

Some students make mistakes with little information or with too many options. As previously mentioned, students need both information and knowledge to make optimal college choices. The college choice model assumes that students make their optimal choices by comparing the benefit-cost tradeoffs between college and major options. Lack of information, misinformation, or unawareness about college options - cost, return, curriculum, major, and special programs - and admissions policies makes it impossible for students to correctly compare the tradeoffs.

On the other hand, when faced with many options, students may also have the overchoice problem, in which they are unable to make good decisions. Students may have limited cognitive capacity and attention when evaluating a large number of choices and identifying the best fit options, for example, identifying a short list of reach, peer, and safety colleges from thousands of colleges with multidimensional information. This problem is exacerbated

[22] The personalized advising process is based on the four modules in the guidebook: (1) the search for college information, (2) understanding admissions policies, (3) equating CEE score and identifying college types, and (4) applying to a set of reach, match, and safety colleges. The guidebook was prepared by an army of experts, including professors and graduate students in the field of both K-12 and higher education policy, school counselors, and college admissions officers. See Chapter I for more descriptions.

in centralized admissions, in that the application period is extremely short, only three to five days.[23] Moreover, students need to apply to colleges and majors simultaneously. They need to compare thousands of college-major options in a short time period.[24]

In the *Bright Future of China Project*, we have designed a structured guide to help students navigate college applications in situations of lack of information, or when facing abundant choices. The personalized advising guides students through three main procedures: (1) providing a comprehensive and reliable guidebook to help them focus on the information that matters most; (2) using data analysis to generate a short list of colleges and majors, and to identify college types based on *ex ante* admission probability; and (3) instruction regarding the steps and strategies to make final decisions.

Some students rely too much on routine. Transitioning from high school to college requires students to deviate from their daily high school routine. During the entire K-12 period, students are taught merely the skills for standardized test, particularly in centralized systems where the college entrance exam score is the sole criteria for college admissions. Nevertheless, right after they take the College Entrance Exam, students have to learn and use a different set of skills to make the complex college-going decisions. Relying too much on routine and automatic thinking may make students less likely to seek new information and thus more likely to ignore better opportunities, even if access to information is free and readily available. In the field, many treated students in the experimental sample favored "a suggested application plan" rather than being taught to make the plan, because the former has a much smaller cognitive cost of changing routine for the students.

The personalized advising program in the *Bright Future of China Project* is designed to help students change routine by nudging a student's learning of the specific knowledge

[23]Most Chinese students only start to think about college choice after they know their CEE scores, about two weeks after they take the CEE.

[24]We have observed many behavioral mistakes of college applications in the field. For example, some students may just choose a college/major based on the college/major name (e.g., they infer a college with the word "national" in its name would be a high-quality institution) and fail to consider school quality and admissions probability. Some students may just choose the in-province colleges without learning about the higher-quality college opportunities in other regions.

for college choice. The "teach a man to fish" principle attempts to help students form some transferable decision-making skills that can be applied to future choices in life. Furthermore, the advising program also helps students review and evaluate their proposed application plans, which could reduce their anxiety about making mistakes. Text message reminders about application tips and deadlines are also used to supplement the main conversations between advisers and students.

Some students focus too much on the present. The use of information and knowledge is associated with immediate and salient cognitive costs. However, the benefit of a deliberate application is uncertain to students. One potentially effective intervention is to offset immediate costs with immediate benefits, such as information about returns to college. However, people may still spend little time making the decisions, despite the complexity and importance of the decision (Benartzi and Thaler, 2007). A large proportion of students complete their college applications too soon to think about the options carefully. During the advising, we used a structured advising process to nudge students to spend enough time working on the application preparation (e.g., the predictions) and to seek more information about each college of interest.

To sum up, the college choice advising program targets the behavioral barriers of students in centralized college admissions systems, in which the information and application process is simplified. Even so, learning and the use of decision-making skills are essential. The experimental evidence in 2016 shows that personalized advising substantially improved low-income students' college access and match by shaping their application strategies. This chapter focuses on expanding the personalized advising program using machine learning predictions and teacher incentive policies.

As will be described in greater detail below, the machine learning intervention attempts to increase advising efficiency by providing machine learning predictions. Automatic predictions reduce a student's learning cost and hands-on workload. Machine learning mirrors the predictions of *ex ante* admission probability that an expert or a sophisticated student would make in college choices, which are the proposed outcomes of conventional human advising. The testable hypothesis is that the machine learning-assisted advising would produce similar treatment effects on college access and match compared with the conventional advising, but with a large improvement in advising efficiency. However, the machine learning predictions do not consider individual preferences, treated students who receive the personalized advising may not change their college choice behaviors given motivated beliefs and strong preferences.[25]

The effectiveness of teacher incentives is ambiguous. Teachers may serve as experienced counselors. However, there are several reasons teachers may be ineffective at helping their students make optimal college choices. Moreover, the current prevalence of undermatch shows that teachers could do better in guiding students. First, regarding the pay-for-performance design, teachers may not be incentivized to provide effective college-going advising if the incentive is too small. Second, even if teachers are incentivized, they may not have sufficient knowledge and only adopt simple heuristics or brief guidelines. Third, even if teachers are knowledgeable about making college choice decisions, they may not guide students in choosing or ranking interested colleges as effectively as the expert advisors do in the personalized advising program. Oreopoulos and Ford (2019) find that the lack of guidance in choosing eligible programs undermines the advising impacts on college enrollment. Lastly, it is possible that advising crowds out a student's intrinsic motivation and so produces negative effects (Damgaard and Nielsen, 2018). With help from teachers, students may rely on the routine of "listening to teachers" and merely follow a teacher's brief or even incorrect guidelines without critical thinking, and without performing the necessary steps to make their optimal decisions.

[25] In the personalized advising by an expert's human instruction in 2016, the expert advisor fully considered each treated student's college preferences (e.g., geographic location, tuition, and major). However, the machine learning-assisted advising in 2017 only provided data predictions regardless of individual preferences in order to increase advising efficiency. The adjustment of application plans based on individual preferences was left for students.

2.4 Interventions

I propose two interventions that scale up the personalized advising program: machine learning-assisted advising and a teacher pay-for-performance policy. During the past decade, in parallel with the increasing use of performance pay policies in schools all over the world, machine learning has been showing its usefulness to behaviorally nudge people in personalized schooling decision-making. Both approaches have the potential to increase school effectiveness in promoting learning and the use of knowledge in college choice.

2.4.1 Intervention for Students: Machine Learning

The first potential scale-up solution is to increase the intensive margin of labor supply. I explore how to provide college-going advising more effectively without changing existing school organizational policies (e.g., hiring more teachers or school counselors), but by changing the ways that advisors facilitate student learning in college choices and applications. Specifically, I examine how machine learning algorithms could improve human college choice decisions.

The 2016 program of the *Bright Future of China Project-Ningxia* shows proof-of-concept evidence that the core of college choice and application, and thus the effective college-going interventions, is the prediction of admission probabilities (see Chapter I). A thoughtful application based on precise predictions substantively improved college admissions outcomes. Students were instructed using guidebook, workshop, or personalized advising about the steps to generate the predictions, after learning about college admissions and strategies for applications. Notably, during the individualized advising program, we assisted students to evaluate *subjective* admissions probabilities through a variety of data analysis tasks, which was both labor-intensive and time-consuming. There is considerable room to improve the efficiency of students or counselors in making predictions and thus to scale up the advising program.

Machine learning algorithms mirror the predictions that a skillful expert or student would make in the college choice process; that is, learning from a history of past admissions outcomes to predict the admission outcome at each college. Subjective predicted probability is the central skill/strategy that we offered to students in the designed college-going guide, particularly in the human instruction approach of one-on-one individualized advising. An expert or a student's subjective prediction is often imprecise, yet still useful when qualitatively correct. For example, we may use a single criterion such as the five-percentile bandwidth to define reach, match and safety colleges (Hoxby and Avery, 2013). Using machine learning, we can calculate the admission probability of each college using multi-dimensional factors such as different measures of admissions scores, quotas, and college/major attributes. Even two safety colleges can differ substantially in their actual *ex ante* admission probabilities. When machine learning excels at precisely predicting *objective* admissions probabilities more than the human instruction approach, why not use it for better decisions?

Machine learning is used as a "black box" methodology that provides the prediction outcomes for the personalized intervention design. I use the unusual complete data on student applications in a whole college matching market, thanks to our long-standing research-practice partnership with the Ningxia Department of Education. I then apply machine learning algorithms to generate precise predictions for each student for each college: notably, each type of students with the same CEE score, gender, and race (gender and race may affect eligibility for special programs or majors). Because of the detailed application data, I can even predict the admission probability for each student at each college-major option of every rank order in the application list.[26]

When a skillful expert, teacher, or student makes her "human predictions," it is difficult to consider many factors simultaneously, such as minimum/maximum/median/mean admis-

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

sions scores of a college in the past few years. All of these "human predictions" are subjective approximations (e.g., reach vs. safety, or unlikely vs. likely). This problem is much simplified in a machine learning algorithm. Similar to Kleinberg *et al.* (2017), I provide "the machine" with a set of student-college-major-order level data from the 2016 cohort, each observation consists of a set of input features, including admissions scores and corresponding ranking, quota, order, student's CEE score, as well as an outcome (admission=1) to be predicted. I then choose a prediction model (algorithm) to minimize the loss function that generates accurate out-of-sample predictions. I use Random Forest, one of the most commonly used supervised learning algorithms, which builds on a collection of decision trees to generate the predictions.[27] Figure 2.9 shows feature importance for predictions. As expected, a student's ranking within the province-track (*paiming*, like an equated CEE score), her normalized CEE score (*normalized_zongfen*), and the rank order of the applications (*zyno*) carry the largest importance shares. A student's subject scores and a college's admissions scores in the previous year are also correlated with the predicted admissions, while other features are not that important. To avoid over-fitting, I randomly partition the data into an 80% training set and a 20% test set. Cross-validation shows that the prediction accuracy is 94.3% (95% confidence interval: 94%, 94.5%; suggesting a high accuracy) with a sensitivity of 83.4% and a specificity of 98%.[28] Since the goal of prediction is college admission, the model is conservative in predicting the admissions among students' reach colleges.

In the field, the intervention provided to treated students was machine learning-assisted online one-on-one advising. The individualized advising program was similar to that in 2016 using WeChat, the most popular message App in China. Two expert advisors were assisted with machine learning predictions and two additional data-assisted steps. First, the advisor

[27]Assessing the performance of different machine learning algorithms vs. human instruction on the same set of students and applications is out of the scope of this chapter. In a companion work in progress, I have tested other models (e.g., a neural network, deep learning) that take much more time to develop. Random Forest overall shows good performance. I recently tested XGBoost, an algorithm that has been dominating applied machine learning, and found that it increased prediction accuracy by about one percentage point.

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

used a program to automatically equate CEE scores for students, which was done manually in 2016. Next, the advisor provided a short list of reach, match, and safety colleges, which aimed to reduce a student's search cost. Eventually, the advisor asked treated students to provide a candidate list of colleges and majors, and the rank orders. The advisor then returned the predicted probabilities of each college-major in the candidate list. Figure 2.2 and Figure 2.3 present examples of the use of data analytics in the personalized advising. Figure 2.10 shows the conversations between the adviser and the student through the online message App.[29] Students picked their final application lists consisting of a group of reach, peer, and safety colleges based on individual preferences and completed their applications. This intervention design combines human judgment and algorithmic judgment in order to minimize any potential errors and to make the best possible college application. During advising, the main task of the experts was to help students check and fine-tune their candidate list. The majority of the conventional human instruction (as in 2016) was then replaced by machine learning in 2017. Therefore, the advising productivity was greatly improved.

I also developed a "business as usual" intervention, serving as a placebo test. I was interested in the following question: In the absence of our interventions, what for-profit consulting services would a student probably have access to? Between 2016 and 2017, I reviewed dozens of Chinese companies that were selling college application consulting services at a price range between 100 RMB and 500 RMB ($15-$80).[30] Excluding the obviously incorrect "application strategies," I kept a brief list of mostly harmless college application advice guidelines, which can be seen as a partial, simplified, and low-touch version of the guidebook that was provided to students in 2016. These tips were then provided to students in the "business as usual" group. A group of research assistants (not expert counselors) also answered general questions. In many cases, students were directed to other online resources for further information. Research assistants did not answer any specific questions about

[29]subsection 2.13.2 provides further descriptions of the personalized advising.

[30]There were also more expensive personalized services charging thousands or tens of thousands RMB. But it is rare for students in the poorest regions to pay that much money. There are also growing "big data," "artificial intelligence" services, but their data and methods (in most cases, incorrect) are not applicable.

individual CEE score equating, short lists of colleges, or application planning.

2.4.2 Intervention for Teachers: Pay for Performance

The second potential scale-up solution is to increase the extensive margin of labor supply by making high school teachers temporary counselors. High school teachers are the primary source of information and assistance that low-income students reach out to for help in their college choices. This is also true in the U.S. (Roderick *et al.*, 2011; Belasco, 2013; Lee *et al.*, 2017). However, teachers may lack incentives and knowledge to provide effective college choice support. Particularly in China, teacher performance is only evaluated by student scores in the College Entrance Examination, but not by their college admissions outcomes. This multi-task problem makes high school teachers not willing to put additional effort toward helping students in the application process. They may also have poor knowledge about college applications.

I designed a pay-for-performance policy to incentivize high school teachers to provide individualized college choice and application assistance to students. This design was motivated by the fieldwork lessons from the first-year program, and was made possible in part by the school organizational characteristics. First, high school teachers in China, especially classroom head teachers, stay with the same class of students for the three consecutive years in grades 10 to 12. As a widespread classroom organization form across countries (Little and Dacus, 1999; Wang *et al.*, 2017), teacher looping enables the head teacher in each class to know her students well, and to become a student's trusted source for college application assistance. Second, since the accountability assessment is solely based on a student's CEE score, teachers of the graduation cohort are off-duty for the entire summer after their students take the CEE. These idled, yet highly educated human resources, may provide an important supply of expert consolers that could scale up the personalized advising program.

One takeaway from the existing literature on performance pay is that the incentive design matters. For instance, Fryer (2013) suggests that the NYC incentive scheme he studies may be too complicated to be effective. We have carefully considered these potential design problems when preparing our pay-for-performance contract. Class head teachers in the incentive program were offered an incentive contract stipulating performance pay based on college admissions outcomes of all students in their classes, conditional on CEE scores. While high school teachers (in our sample, about 50%) may receive some form of performance pay based on CEE scores, there are no monetary or non-monetary rewards - except intrinsic motivation - for their work in advising students in college choices and applications. Collaborating with the Ningxia Department of Education, and funded by research grants, we added a new and clearly defined performance pay scheme for class head teachers based on class-level performance in college admissions.

Class-level "value-added" is measured using the admitted college median CEE scores of all the students in one class. College median score is more straightforward than the factor index of college match, the primary outcome used in this chapter and in Chapter I. I use the "pay for percentile" approach because it induces teachers to allocate socially optimal levels of effort to all students in the class (Barlevy and Neal, 2012).[31] The percentile performance measurement proceeds as follows. Students are placed in comparison groups based on their CEE score ranking percentiles. Within each of these comparison groups, each student is then assigned a rank percentile score based on her college choice outcome (i.e., college peer median CEE score). Students who do not enroll in college will be assigned the lowest ranking. A teacher's "value-added" measure is the average percentile rank in her class.[32]

I then rank all the class head teachers (about 1,100) in Ningxia using the class-level average percentile rank. The incentive contract is structured as:

[31] Loyalka *et al.* (2019) present experimental evidence that Chinese teachers offered "pay for percentile" incentives outperform teachers offered simpler schemes based on class average achievement or average gains over a school year. But Gilligan *et al.* (2018) show that teachers in Uganda responded to a pay for percentile incentive system in ways that raised attendance rates only in schools where textbooks were provided.

[32] One small implementation issue is that we mentioned to the treated teachers that we would create this ranking percentile measure, but did not explain it in detail. Although we cannot rule out the possibility, this should not strongly impact the intervention effects. This will be discussed later.

$$Teacher\ Performance\ Pay = \begin{cases} 5,000\ RMB, & \text{if } Rank \geq 95th\ percentile; \\ 3,000\ RMB, & \text{if } Rank \geq 70th\ percentile; \\ 0, & otherwise. \end{cases}$$

I considered several concerns about performance pay policy raised by Fryer (2013). First, *are the incentives not large enough*? The average monthly salary for high school teachers is about 2,000-3,000 RMB in our experimental province. Our teacher survey data confirm this.[33] Note that the college application period is in total no more than two weeks. Then the expected bonus payment is as much as their monthly salaries - about 8 percent of annual salary, and more than many of the performance pay plans in the existing literature. I also collaborated with the Ningxia Department of Education to make the contract credible and to add non-monetary incentives (award plaques) to eligible teachers. Second, *is the incentive scheme too complex?* I used the "pay for percentile" approach and informed teachers how we would evaluate their relative whole-class performance. The research assistants contacted each treated teacher to make sure that all of the treated teachers understand the pay scheme. Third, while group-based incentives may not be effective due to the free-riding problem, we used an individual-level incentive plan that each treated teacher would only be evaluated by the college admissions results of students in her own class. Lastly, *do teachers know how they can improve college applications*? As will be investigated later, I argue that this is one likely reason that this performance pay policy may be ineffective.[34]

To address the last concern that teachers may lack sufficient knowledge of college applications, online training sessions were also provided to treated teachers, but not to control

[33]Before the fieldwork, we conducted interviews with high school principals and teachers. Most of them reported that a bonus payment of 1,000 RMB seemed attractive to them.

[34]In the pre-fieldwork interviews, teachers reported that in past years many students reached out to them for assistance and guidance in their college applications every year. Teachers often felt poorly prepared to provide students with detailed guidance. This did not surprise us. It is the reason why we attempted to design the teacher intervention.

teachers, before the college application period. Similar to the school workshop in 2016 as described in Chapter I, I provided information about the prevalence and causes of the undermatch problem. Teachers were recommended of various school-based approaches to help students in the college application process such as workshops, collective application, parental engagement, information about college fairs, and individualized advising. Additionally, treated teachers were encouraged to attend online mini-lectures about the key steps for one-on-one advising, which aimed to ensure that all teachers would provide the same "human instruction."

2.5 Experimental Design

2.5.1 Data Sources and Motivating Evidence

I collected and analyze unique, large-scale student-level administrative data for the universe of the 2017 high school graduation cohort in Ningxia province. The data was provided by the Ningxia Department of Education and the Ningxia Education Examination Board, the provincial centralized administration office of the College Entrance Exam and college admissions. Using accurate administrative data to analyze the entire population of applicants in a college matching market (not a sample), and without problems of missing data and sample attribution, this chapter can identify students' college application behaviors (strategies and preferences), admissions outcomes, and enrollment decisions.

The confidential student-level data include student demographics and high school attendance records, CEE scores, full rank-order applications data, and admissions results, which are the same as I used for the 2016 graduation cohort in Chapter I. I discovered and cleaned a new dataset of the college application submission time, which will be used in analyzing the intervention's effects on the timing of applications. I also constructed detailed student-college-major-choice level "big data" (over 1 billion observations) to develop the machine learning algorithms that predict the admission probability of every college-major option in every possible application rank order for every student.

I merged the college-major level information, which consists of address, tuition, quota, prior-year admissions scores, with the student-level data in order to study their college choice strategies and preferences. The college-major level information was the same as that provided to students during the college application period by the Ningxia Department of Education in printed books.[35] I merged in additional college-major data, including elite college identifiers and national college rankings. Combining student-level data and college-major data, I created a series of variables measuring college match outcomes and college application behaviors. Finally, I extracted the track-tier admissions cutoff scores from the Ningxia Department of Education's official website. Additionally, data from teacher surveys - administered by the Ningxia Department of Education before and after college applications - were used for the teacher-level randomization during the teacher pay-for-performance incentive interventions.

Motivating evidence of college undermatch. As discussed in greater detail in Chapter I, low-income students, using rural *hukou* (household registration, the primary source of income inequality in China) as a proxy for poverty, are much more likely to undermatch compared with their higher income peers. This chapter finds qualitatively the same results using data from the 2017 graduation cohort with the 2016 cohort as discussed in detail in Chapter I. Figure 2.1 presents the distribution of substantial undermatch of both rural and urban students. Undermatch is conservatively defined as follows: students undermatch when they are admitted to a college with peer median CEE score 0.25 standard deviations lower than their own CEE score, or they are not admitted to any colleges. High-achieving, low-income students have a particularly high undermatch rate. In centralized admissions, the between-group difference in college undermatch is solely due to college choice behaviors when controlling for CEE scores.

Table 2.2 replicates the descriptive analysis of the correlation between college choices and

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

admissions outcomes as discussed in Chapter I. Using the sample of students who submitted their college applications and were not in the treatment groups in 2017, I first identify the poverty gap in college match using the following linear model, which I estimate by OLS:

$$Y_i = \beta_0 + \beta_1 * Rural_i + \gamma * X_i + \varepsilon_i \tag{2.1}$$

where β_1 measures the rural-urban gap in the college match outcomes Y_i, holding individual covariates X_i (CEE score, demographics) equal. The outcome variable Y_i, which will be discussed later, is a principal-component index of five continuous college quality measures, including college admissions scores in Ningxia in 2017 (median, mean, minimum) and national college quality measures (standardized score and ranking percentile).[36]

Column (3) shows that even having the same CEE scores and demographic characteristics, rural students are admitted to a college with 0.107 standard deviations ($p<0.001$) lower quality than urban students.[37] Since students in the same high school may share the information and support from the school, controlling for high school fixed effects reduces more than one-third of the gap (β_1=-0.063 in column 7). However, the sizable rural-urban gap in college match persists, and is attributable to their different college choice behaviors.[38]

I next examine the correlation between college choice behaviors (strategies and preferences) and admissions outcomes. Following Chapter I, I constructed a series measures of strategies and preferences using students' full applications data. Appendix subsection 2.13.1 provides a detailed description of these measures. I add the six principal-component indices (standardized) stepwise in Model (2.1). The data-based targeting strategy is the core of

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

[37] Table 2.10 shows that, including those who did not apply to any college (and were assigned the track-tier lowest admission score), the raw rural-urban gap in the college match index is -0.134 standard deviations. Rural students end up with colleges on average 0.149 s.d. lower quality when controlling for CEE score and demographics. Female and minority students are also more likely to undermatch, while repeaters are better matched than first-time exam takers.

[38] All the high schools, located in urban districts, have a mix body of students from rural or urban families. Controlling for class fixed effects or neighborhood fixed effects does not change the results once we control for school effects. Using other college access and match measures as discussed in Chapter I shows similar poverty gaps.

college application knowledge as well as the designed guide (and the personalized advising), which enables students to apply for a targeted set of reach, peer/match, and safety colleges.

The first two columns of Table 2.2 show the sample means of each college choice behavior index. Rural students are less likely than urban students to use the targeting strategy and follow the general advice (e.g., fill in all the major applications within each college). Moreover, rural students prefer colleges with lower tuitions and larger admissions quotas, and in-province colleges that would limit other high-quality college opportunities (Hillman, 2016; Ovink *et al.*, 2018). Notably, column (4) shows that a one standard deviation increase in the use of the targeting strategies is significantly correlated with a 0.2 s.d. increase in the quality of the admitted college. The correlation is stable when we control for other behavior measures in column (6). The other strategies and preferences are also correlated with admissions outcomes, but in smaller magnitude. Comparing the changes in the rural-urban gap (β_1), we find clear evidence that targeting strategies explain the largest proportion of the variation in the outcome (40 percent), controlling for CEE score and demographics. This result remains unchanged when we control for school fixed effects in column (10). Targeting strategies explain about half of the rural-urban gap in college match.[39] Overall, as displayed in columns (6) and (10), if rural students have the same college choice behaviors as urban students, the rural-urban gap reduces by more than 60 percent.[40]

The results confirm the previous findings in Chapter I. Academic undermatch is prevalent in centralized admissions systems as well. Disadvantaged students are more likely to undermatch, even when they have the same CEE scores. Moreover, the data-based targeting strategies for applying to a set of reach, peer, and safety colleges are highly correlated with improved college match. The college application guide interventions in 2016 (guidebook, workshop, personalized advising) show that the "human instruction" approach accompany-

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

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

ing these strategies improves students' college choice behaviors and admissions outcomes. I now turn to the randomized experimental designs to examine the effectiveness of the two proposed scale-up solutions, both of which focus on improving the individualized advising of college choice knowledge and strategies.

2.5.2 Experimental Design: Individual Level Randomization

Following the first-year program in 2016, in 2017 I conducted another set of randomized experiments in Ningxia. Table 2.1 summarizes the experimental design. Figure 2.7 shows the timeline. To increase statistical power, and having gained increased trust from local government and school leaders, I used individual-level (students and teachers respectively) stratified randomization. This was possible because I was able to communicate with students and teachers through the official channels provided by the Ningxia Department of Education. I focused on both rural and urban students, since most of the students in Ningxia (one of the poorest provinces in China) are from relatively low-income families and lack guidance in college choices.

Student Interventions

The student randomization sample includes the universe of high school graduates who took the CEE in 2017. I implemented stratified student-level randomization. Using student information from the College Entrance Exam Registration Data, I generated randomization strata by school, track, gender, race, rural *hukou*, county of residence, and achievement. For before-CEE randomization I classified high-achieving students using low-stakes high graduation test scores. Students who ranked above the 75th percentile in the high school graduation test - held in the fall semester of senior year - were classified as high-achieving students.

Students were randomly assigned into one of the three groups: (1) 5,647 students were provided access to the machine learning assisted personalized advising; (2) 5,370 students were provided access to a generic version of the business as usual for-profit consulting service; and (3) the remaining 43,038 students served as the control group. Since targeting strategies may be more useful for high-achieving students (Tier 3 and Tier 4 colleges are not selective), I disproportionately assigned more high-achieving students (classified using graduation test scores) to the machine learning group (45%; 28% in the "business as usual" group).

Implementation. When the CEE score report became available online, and students could start to apply to college, the Ningxia Department of Education sent out text message reminders to every student using the cellphone numbers from students' CEE registration records.[41] Immediately after that message, the Ningxia Department of Education sent another message to students in the two treatment groups, introducing the individualized advising opportunities provided by experts from Peking University and Ningxia University. Students were encouraged to contact us using the online chat App. The text message was the same for the two intervention groups, except for the contact information.[42] The research assistants verified each student's ID information and asked each student to complete a short online survey. After that, students were directed to either a "machine learning" individual chat group or a "business as usual" individual chat group. Each chat group had three members: the treated student, one of the two expert advisors (one research assistant in the "business as usual" group), and an administrative assistant.

The "machine learning" intervention proceeded using three data analysis steps as described in Section 2.4.1. Students were provided with (1) automatic CEE score equating, (2) a short list of colleges, and (3) the predicted admissions probabilities of their candidate lists of colleges and majors for their proposed rank orders. The main task of the expert advisor was to help students check and fine-tune their candidate lists, and to provide more detailed guidance on specific questions. Most students kept in touch with inquiries and questions

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

[42] For better coordination, we had five user account numbers, three for the machine learning group and two for the "business as usual" group.

until they submitted their applications, and many of them informed us of their admissions results in July and August.

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

Summary statistics and validity. Table 2.11 shows that student characteristics are well-balanced across groups given the student-level randomization. Student populations are very similar in the 2016 and 2017 cohorts. About 57% of students are from rural families; 32% are minorities; and more than 23% have repeated the 12th grade at least once. Because the randomization for the student intervention was independent of that for the teacher intervention, there were 836 treated students in treated teachers' classes. Balance checks are still valid when we exclude those students. Table 2.12 confirms that student characteristics have statistically zero prediction power for the treatment status. All the joint F tests are statistically insignificant.

Teacher Intervention

I recruited a sample of class head teachers in the graduation cohort of 2017, using a teacher survey before the College Entrance Exam. The survey was prepared by my research team through the online survey platform at Peking University and was formally organized by the Ningxia Department of Education. The survey was used to collect background and contact information, and to inform teachers of the importance of college application as well as the credibility of possible future contacts from the research team. Teachers did not know anything about the incentive policy before the treated teachers received the formal contract letter.

Of about 1,100 head teachers for the 2017 graduation cohort, 267 responses were received between June 1 and June 5, during which head teachers were almost off-duty because the CEE started on June 7. Teachers without valid contact information, those with many careless answers, and those who were not able to be matched with the records in students' CEE registration data, were dropped. The final sample consisted of 184 teachers, whose classes contained 9,354 students with an average class size of 51. I created strata using a school-track-elite class classification.[43] I randomly assigned each teacher to either the treatment group - which received the pay-for-performance contract and contained 88 teachers - or the control group, which had 96 teachers.

Implementation. In the teacher survey, teachers were informed by the Ningxia Department of Education that a group of researchers from Peking University and Ningxia University would directly contact them to provide additional assistance regarding students' college applications (see Figure A of Appendix Figure 2.12). Using teachers' contact information from the survey, I sent out an information letter to teachers who were randomly assigned to the treatment group (see Figure B of Appendix Figure 2.12). The information letter was sent to each teacher's email address on 20 June in the name of the China Center for Education and Human Resources Research at Peking University. The research assistants then contacted each teacher to make sure they received and read the letter. The letter provided a brief introduction to the undermatch problem and how high school teachers could help, and nudged teachers to provide personalized advising to their students. Treated teachers were also informed about two online live training sessions on the 22nd and 23rd of June and that were particularly prepared to assist them in providing effective support to their

[43] I aimed to create as many strata as possible with the goal of having balanced groups within each stratum including at least two teachers per group (Athey and Imbens, 2017). If the initial strata had more than 8 teachers, we further differentiated the strata by school-track-elite-teacher experience (longer than 10 years). If there were fewer than 4 teachers, we combined these strata by county-track-elite.

students.[44]

On the morning of June 24, when selective college eligible students started to apply to college, I sent the pay-for-performance contract letter to the treated teachers (see Figure 2.13). The timing was carefully chosen after many conversations with district and school leaders. Teachers and students started to work on the college applications only after students could apply to college.[45] I also wanted to amplify the treatment effects by giving teachers a timely and pleasant surprise. The letter introduced the performance pay scheme to treated teachers. To ensure the credibility of this policy intervention, the letter stated that the contract was funded by a Chinese Natural Science Foundation grant and also presented the official contact information. The research assistants contacted each teacher again to make sure that all the treated teachers understood the performance pay contract and agreed to opt into the contract. In August 2017, after all the students applied to college and received admissions results, I conducted a brief follow-up survey among all the 184 teachers in our experimental sample, in both the treatment and control groups.

Summary statistics and validity. Table 2.13 shows the teacher-level characteristics from the pre-treatment survey. The treatment and control groups are well balanced, and the joint F-test results in a 0.478 p-value. Class head teachers are experienced in teaching with an average of 13 years of teaching experience, more than half of them have advanced teaching certificates, and about half of them report professional proficiency in English. Those who report their monthly salaries earn an average of about 3,500 RMB ($500). Around 30 percent of teachers do not reveal their salary information. Fewer than half of these teachers agreed or strongly agreed with the statement that "college applications impact students' college ad-

[44]We informed teachers that our two co-Principal Investigators would give the training sessions and what would be covered in the sessions. We also provided three contact methods for them to stay in touch. The online training lecture video is now available at https://www.bilibili.com/video/av11582407/.

[45]Students got to know their CEE scores and could start to apply to selective colleges during the late afternoon of June 23. Head teachers were busy distributing CEE score reports and doing other administrative tasks on that day. Without other school policies (e.g., accountability), it was not feasible to provide teacher training sessions before the CEE. Teachers also showed little interests of attending such training sessions between students took the CEE and students started to apply to college.

missions outcomes."[46] The student-level balance check results in Table 2.14 present similar results, and show that students in both groups have on average statistically equivalent characteristics.[47] Moreover, students in our experimental sample look similar (representative) to those not in the experimental sample. One exception is that there are more repeaters, and thus on average higher achieving students in the non-experimental sample. It is difficult to identify the class head teachers for repeaters because they may have reported their first-time 12th-grade teachers rather than their teachers when they repeated the 12th grade.

2.5.3 Econometrics

Effects of student interventions. To estimate the causal impact of the student interventions - "machine learning" and "business as usual" - on outcomes, I estimate an intent-to-treat effects (ITT) regression:

$$Y_i = \beta_0 + \beta_1 * T1(Machine\ learning)_i + \beta_2 * T2(Business\ as\ usual)_i + X_i * \gamma + \delta_s + \varepsilon_i \quad (2.2)$$

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

X_i includes a student's CEE score to identify the "college choice" effect. In centralized systems, college admissions are jointly determined by a student's entrance exam score and her choice. Controlling for the entrance exam score in X_i, β_1 and β_2 estimates the treatment effects on a student's college access and match, through the impacts on her college choices. While treatment and control groups are well balanced in student demographics because of

[46] Another 25 percent chose "somewhat agree." When asked a similar question in the post-treatment survey, more than 85 percent of teachers agreed that "a good application could bring students better admissions outcomes."

[47] Table 2.15 presents the OLS regression results predicting the treatment status, from which we perform the joint F test. Teachers are balanced in both teacher-level and student-level characteristics.

the stratified randomization (as shown in Table 2.11 and Table 2.12), X_i also controls for a vector demographics (gender, race, age, rural, track, repeater from the prior years) in the preferred specification to increase the statistical power and to reduce potential biasedness of effect size estimation. As expected, the results do not change if I exclude those student-level covariates.

Model (2.2) identifies the impacts of *being offered access to receive* personalized advising in college choice and application. I also estimate the treatment-on-the-treated effects (TOT) using a 2SLS regression, which measure the average effect of *receiving* the personalized advising on those who actually receive it. The first-stage regression examines the take-up of the two advising interventions:

$$\begin{aligned} Treated\ in\ T1_i =& \beta_0 + \beta_1 * T1_i + \beta_2 * T2_i + X_i * \gamma + \delta_s + \varepsilon_i \\ Treated\ in\ T2_i =& \beta_0 + \beta_1 * T1_i + \beta_2 * T2_i + X_i * \gamma + \delta_s + \varepsilon_i \end{aligned} \tag{2.3}$$

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

$$Y_i = \beta_0 + \beta_1 * \widehat{Treated\ in}\ T1_i + \beta_2 * \widehat{Treated\ in}\ T2_i + X_i * \gamma + \delta_s + \varepsilon_i \tag{2.4}$$

where the other specification issues are the same as in Model (2.2). ITT effects show the overall effects that we could expect if the program implementation is similar to what we did in 2017 when the take-up was low. TOT effects identify the potential intervention effects if we provide the machine learning assisted advising with stable program productivity to all, and all students take up their opportunities.[48]

I examine the primary college admissions outcomes as will be described in the following section, which include both extensive and intensive margins, because the interventions are primarily designed to improve college access and match. See Chapter I for further discussion of the outcome measures. I also examine a list of exploratory measures of college choice

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

behaviors. These measures are from the same domain and highly correlate with each other, so that the multiple hypothesis testing problem is minimal. Furthermore, I use the aggregated indices of each group of outcomes from a principal-component analysis. Additionally, I apply the method proposed by List *et al.* (2016) to check the robustness of the results.

Effects of teacher intervention. In order to evaluate the effectiveness of teacher incentives, I estimate a similar intent-to-treat effects (ITT) regression as in Model (2.2):

$$Y_{ic} = \beta_0 + \beta_1 * Pay\text{-}for\text{-}Performance_c + X_i * \gamma + \delta_s + \varepsilon_{ic} \tag{2.5}$$

where Y_i is the outcome of interest for student i in teacher c's class. $Pay\text{-}for\text{-}Performance_c$ is an indicator variable for teacher c receiving the pay-for-performance contract. δ_s are strata fixed effects. X_i includes the student's CEE score and demographic characteristics. Results do not change when adding teacher-level pre-treatment covariates, except for that the standard errors decrease a tiny bit. All standard errors are clustered at the school level. I also report standard errors clustered at the teacher level. Since the take-up of the teacher incentive contract was 100%, the TOT and ITT are identical.[49]

2.6 Results

2.6.1 Effects of Machine Learning Assisted Advising

Take-up (first-stage results). Table 2.3 reports the first stage regression results from Model (2.3), separately for the two student interventions: machine learning assisted advising and "business as usual" advising. Column (1) shows that on average, 3.6 percent of students (210 out of 5,647) who were provided the machine learning advising invitations eventually received the personalized assistance.[50] Column (5) shows that on average, 2.4 percent of students (134 out of 5,370) who were provided the "business as usual" advising invitations

[49]We could potentially estimate the TOT of teacher advising behaviors, using the random assignment of incentives as an instrumental variable. However, as discussed in the following section, the pay-for-performance plan did not impact teacher behaviors.

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

eventually received the low-touch personalized assistance. As expected, the results do not change when controlling for student-level covariates or the two random assignment indicators ($T1$ and $T2$) simultaneously. Columns (3) and (7) show that teacher incentives do not impact the take-up rate. High-achieving students had slightly higher take-up (4.93% in $T1$ and 4.14% in $T2$; Chow test p-value = 0.209, which indicates no statistically significantly different take-up between the two groups). F statistics reject the null that the random assignment is a weak instrument for the actual take-up.

The take-up rate is somewhat surprisingly low, but higher than that in 2016. When designing the interventions, we planned for an expected take-up rate of 20% and prepared accordingly a team of expert advisers; thus we did not provide the conventional expert "human instruction" in 2017. One reason for this low take-up is the verification process (Alatas *et al.*, 2016).[51] For the research purpose of identifying the treated students, student exam IDs and school IDs were used to verify and screen the targeted students. That is, I purposely denied most of the would-be always-takers from the control group. We also asked treated students to complete a 10-minute survey before the expert advisor provided advising services. As Hoxby and Turner (2013) suggest, students and parents may often be suspicious of this verification process, even though we only asked for their exam ID and school ID that could not be personally identified without the administrative data from the Department of Education. Over 1,800 users (students or parents) added the contact accounts as friends, accounting for 16 percent of the randomly assigned treated students (assuming few non-compliers). However, we finally provided college application advising to 347 high school graduates in Ningxia in the 2017 program. While this low take-up does not impact the internal validity of the estimates, it results in limited statistical power (e.g., significance tests among high achieving students) and an inability to infer the intervention effects and

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

effect heterogeneities on a large scale.

Did machine learning work? I examine the same set of college access and match outcomes as explored in Chapter I. I find that the personalized advising program with the assistance of machine learning and related data analytics substantially improved college access - admission to a college, and match - the quality of an admitted college. Table 2.4 reports both the intent-to-treat effects (ITT) and the treatment-on-the-treated effects (TOT).

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

On the intensive margin, the personalized advising also improved college match, which is measured by the quality of a student's admitted college. Column (5) of Table 2.4 examines the impacts on the college match index, which summarizes a family of five college quality measures using factor analysis. On average, treated students were admitted to colleges with 0.598 standard deviations higher quality as measured by the single index, statistically significant at the 0.05 level. The corresponding ITT effect of providing access to personalized advising is 0.022 s.d. ($p<0.05$). Given that college admissions are solely determined by a student's CEE score and her applications, the results suggest that in the counterfactual situation without receiving the machine learning assisted advising, a treated student had to increase her CEE score by 0.598 s.d. to be admitted to the same college. Comparing with many possible inputs in K-12 education, this chapter presents a very effective, and relatively low cost, behaviorally-designed intervention to improve college access and match for low-income students.

The college match index consists of both contemporaneous college admissions scores (median, mean, minimum) and static (in the short term) national college ranking measures (standardized score and ranking percentile). The national college quality measures were constructed using college admissions data in all provinces from 1996-2017, as well as administrative data on institutional resources for every college in China. I use the national measures to minimize the potential bias of using admissions results from the same within-province cohort to denote college quality, for example, a college with few admission quotas occasionally admitting high-achieving students. The static measures also enable us to compare estimates across years.[52]

In column (6) of Table 2.4, I examine the impacts on the national college quality measure. Results are similar in that the personalized advising assisted the treated students to be admitted to colleges with a 0.77 s.d. higher quality in the national college ranking. Table 2.16 presents very consistent results using the other four itemized college match outcomes. Column (7) excludes about 10,000 students who were not admitted to any college and presents the underestimated effects of "machine learning" (overestimates for "business as usual" effects). The point estimates remain large (TOT=0.262 s.d.), but are not precise enough to be statistically significant. Lastly, column (8) uses the conservative dichotomous measure of undermatch as described in Figure 2.1, and shows that the machine learning advising program reduced undermatch by 28.8 percentage points. In magnitude, this equals

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

the control group mean, or more than twice of the rural-urban gap.

Why did "business as usual" not work? In stark contrast, the business as usual advising, as a placebo test that mimics a typical for-profit - and perhaps non-expert - consulting service, had a zero or even negative impact on college admissions. For the primary outcomes of college access and match (columns (1) and (5)), the joint equality tests reject the null hypothesis that the two individualized advising interventions had the same effects. This finding is consistent with the findings in Oreopoulos and Ford (2019), that decreased guidance in choosing eligible programs would limit the effectiveness of advising programs. One reason for the possible negative impact of "business as causal" intervention is that students may make their repeating choices based on general advice such as "repeating increases college opportunities." In contrast, the machine learning-assisted advising nudged students to consider all the possible good college opportunities before deciding to repeat[53] This argument is supported by the results in Table 2.4. The "machine learning" advising largely decreased repeating and accordingly increased on-time college enrollment. The "business as usual" had the opposite effect. Students who received the "business as usual" advising were more likely to repeat the 12th grade and less likely to enroll at a college in 2017. Excluding those who were not admitted to any college, column (7) shows that "business as usual" advising had a small and insignificant positive effect on college match. However, this estimate was biased upward.

How did "machine learning" work? I designed the college application guide and advising to improve a student's college choice behaviors. Using the unique data on students' full college application lists, I test whether the improved college admissions outcomes stem from their application behavior changes. The construction of the (partial) strategy and preference measures is discussed in detail in Appendix subsection 2.13.1, adopted from

[53] It is arguable whether repeating is a good strategy. Goodman *et al.* (2018) show that retaking the SAT improves admissions-relevant SAT scores. But Chinese students have to spend a whole year before retaking the CEE. In this chapter, we define the not on-time enrollment as undermatch because too many students make repeating decisions without thoughtful considerations about college choices. This means that, presumably for some students it is optimal to retake the CEE, but the number of students who actually retake the CEE is much larger.

Chapter I. I find compelling evidence that the machine learning advising nudged the treated students to use the correct data-based strategies to apply for match colleges.

Table 2.5 reports both ITT and TOT effects on college choice behaviors. Column (1) shows that, students who received the machine learning based advising were 30.6 percentage points ($p<0.05$) more likely to apply to at least three colleges in the recommendation list, compared with that 30.5 percent of the control group students who applied to at least three colleges in the list. The list includes all the colleges that we ever recommended to any treated students with an estimated admission probability larger than 35%. This result clearly demonstrates that treated students followed our advising. In contrast, "business as usual" advising did not cause students to be more likely to apply to colleges in the list.

Column (2) uses the single principal-component factor index to summarize the effects on college choice behaviors. The machine learning advising largely and significantly (at the 0.1 level) impacted students' college choices. The next six columns present detailed results for each strategy and preference category (factor indices from a series of items). Results show that the effects of machine learning advising concentrated on improving students' targeting and general nudge strategies in their college choices. This finding is very consistent with the descriptive results in Table 2.2 that these two groups of strategies are the most important factors driving college match, and is also consistent with the focus of the individualized advising intervention. In Table 2.17, we report the itemized results for the targeting and general nudge strategies. Results show that the improvement was not in a few occasional measures, but was universal across the domain of "good applications."[54] As a placebo test, the "business as usual" intervention shaped students' applications in the opposite direction: The treated students were less likely to use the optimal strategies based on data analytics.

Neither the "machine learning" nor the "business as usual" interventions impacted student preferences.[55] This is different from the guidebook and workshop interventions in 2016,

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

[55]Special programs can be seen as an application strategy that students may gain access to higher quality

which shifted students from in-province colleges to out-of-province colleges. Individualized advising did not change student preferences regarding geographic locations. The main reason is that we only assisted students with making optimal decisions given their preferences. In 2016, considering out-of-province opportunities was listed as an application nudge in the guidebook and workshop interventions.[56] The results suggest that the college choice individualized advising using machine learning predictions worked through improving application strategies without impacting individual preferences for colleges and majors.

Treatment effect heterogeneity by income, gender, race, and achievement. I investigate whether the ITT effects on college choice behaviors and admission outcomes varied by income, gender, race and initial academic achievement, using a linear interaction model (Table 2.18). The second row suggests that there were no heterogeneous take-up rates between groups except that high achieving students were more likely to participate in the advising program. Therefore, the difference in the ITT effects is similar to that in the TOT effects. Results show that rural, male, non-minority, and low-achieving students benefited more from the machine learning-assisted personalized advising. Consist with the explanation of the impact mechanisms, those students were more likely to follow the experts' suggestions by applying to colleges on the recommendation list, and to use good college choice strategies.

Comparing the effectiveness of machine learning and human instruction. I have shown robust evidence that machine learning-assisted advising substantially improves college access and match outcomes, through the combination of human expert instruction and machine learning predictions. A final analysis is to compare the effects of the "machine learning" approach in 2017 with the expert "human instruction" approach in 2016 in Chapter I. Using the estimated effects of the expert human instruction in 2016 as a benchmark, I find that the machine learning approach had a similar impact on students' college access

college if they choose to or be able to meet the special program restrictions (e.g., special majors or affirmative action programs). They can also been seen as student preferences for college quality over those special program restrictions.

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

and match outcomes. For instance, the estimated effect of human instruction in 2016 on the college match outcome index is 0.210 s.d. for high achieving students. The machine learning assisted advising generates a 0.285 s.d. effect for high achieving students and a 0.598 s.d. for all students. Using national college ranking as a constant measure of college quality across years, both the machine learning approach and the expert human instruction approach largely shifted students to higher ranked colleges (by about 10 to 20 percentiles).

While the intervention effect was similar using either conventional expert advising or the machine learning assisted advising, machine learning greatly replaced human labors. There were six expert advisers who worked relentlessly to help 119 students in 2016 . But in 2017, only two expert advisors served 213 students. A simple calculation will show that, with the assistance of machine learning in simplifying the data analysis process, the efficiency and productivity of human instruction in the personalized advising program dramatically increased.[57] If we incorporate the administrative process into an online automatic system (e.g., artificial intelligence, Page and Gehlbach, 2017), the need for administrative assistants will be largely reduced (even to zero). To precisely quantify the increased productivity due to the introduction of machine learning remains as an open question for future research

2.6.2 Thinking Fast? Slow!

Most of the existing literature on college-going emphasizes nudging students to take required actions on time to meet the application deadlines. I turn to the other aspect of time use in decision-making: haste makes waste. The seminal work by Kahneman (2011) describes two systems of human thinking - System 1 (thinking fast), and System 2 (thinking slow). System 1 forms automatic, first impressions of decision-making without deliberation, while System 2 involves problem-solving, analytical, and critical thinking.[58] People generally

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

[58]This classification is similar to our information/knowledge classification as described in detail in Chapter I. Acquiring information engages System 1 thinking, but the use of knowledge requires more detailed and

think that they do most of their decision-making using System 2 thinking, making rational and optimal choices. However, they often go with System 1 thinking because of cognitive ease.

As discussed in the theoretical framework of Section 2.3, high school graduates may rely on their daily routine of System 1 thinking. The primary task of preparing for the college entrance exam, particularly in China, is to help students practice as much as possible and to train them for fast thinking during the test. Behavioral problems arise when students make their college choices without thinking enough about their college options, resulting in not making a good college choice. The newly available data on the exact date and time of all student application submissions in 2017 enables me to examine how the length of decision/thinking time correlates with college access and match and how the behavioral interventions nudged students to slow down.[59]

Using the sample of untreated students, Figure 2.4 shows the distribution of application time for students who were eligible to apply for selective colleges in 2017. While most students were patient enough to submit their applications on the last day, a large number of students submitted their applications much earlier than the deadline. Column (1) of Table 2.19 shows that urban students on average spent 61 hours (2.5 days out of 5 days) finalizing their college choice plans after knowing their CEE scores. Rural students were 5.6 hours quicker than urban students to submit their applications. Column (6) shows that about 59% of urban students spent at least two days and that rural students had a 5 percentage points lower probability of doing so. In column (2), we excluded those who did not apply (coded 0 hours in column 1), and there was still a 3.6-hour gap between rural and urban students. The gap reduces to 2.3 hours after controlling for CEE score and individual demographics.[60] Neighborhood and school fixed effects (separately) explain about a 1-hour

specific problem-solving in System 2.

[59]The analysis of college applications and admissions in the previous subsection is pre-specified in Chapter I of the 2016 program. The discovery of the application time data is unexpected and the analysis in this subsection is exploratory.

[60]Students who were high-achieving, in STEM track, a repeater from prior years, and older than 18 years, spent more time in college choices. Female students spent less time. There was no difference by race.

difference between rural and urban students.

Later submissions of college applications do not necessarily imply that students spend more time in making their decisions, because they may finalize their plans early but delay the submissions (i.e., procrastination). Nor does it imply that students make better college choices. I cannot directly test the first possibility. For the latter, in Table 2.20, I examine the correlations between the time of submission and college choices. For each outcome, I report estimates from four different strategies: (1) OLS without school fixed effects; (2) OLS with school fixed effects; (3) inverse-probability-weighted regression adjustment; and (4) IV using the random assignment to the two advising interventions as instrumental variables.[61] Results show that spending more time making college choices is strongly positively correlated with better college access and match. The accordingly improved college application behaviors, especially in data-based targeting strategies, suggest that the "slow" students have made a conscious effort to utilize information and strategies.

The behavioral intervention design in this chapter aims to promote students' System 2 thinking as well as effortful data-based predictions and targeting strategies. The conceptual framework predicts that the treated students should slow down and spend more time carefully thinking about their college choices. In Table 2.6, I test whether the individualized advising slowed down students. The first column shows the average ITT effects. Machine learning, on average, increased application time by 1 hour. The rescaled TOT effect is 27 hours ($p<0.05$). The effects concentrate on rural, male, non-minority, and low-achieving students. This finding is consistent with the results in Table 2.18 that those students benefited from the machine learning assisted advising more than others. In contrast, I find that the "business as usual" intervention decreased students' decision time (ITT=-0.777; TOT=32), which is consistent with the suggestive explanation that the brief guidelines in "business as usual" may have reduced the intrinsic motivation of students, and may have made them less likely

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

to form thoughtful college choices.

2.6.3 Effects of Teacher Incentives

To evaluate the impacts of teacher incentives, I focus on the same set of college admission outcomes as examined in the student interventions. The results are not encouraging. Table 2.7 shows that paying additional bonus rewards to teachers has a statistically zero impact on all the students' college access and match outcomes. If anything, teacher incentives may increase college applications by 0.4 pp and college admissions by 0.2 pp on average. The coefficients are precisely estimated zeros. Column (8) shows that students in the classes of treated teachers were about 1.5 percentage points more likely to be admitted by an undermatched college, but the impact was not statistically significant. Consistently, teacher incentives have negative, but close to zero, impacts on the quality of colleges that their students were admitted to.[62] The results do not change when we control for teacher covariates or cluster the standard errors at the teacher level (rather than at the school level).

Why did teacher incentives fail? After college admissions ended in August 2017, I conducted a short post-treatment survey among all the 184 teachers in the experimental sample. Table 2.8 compares the group mean differences in belief, knowledge, and action of college application advising between the control teachers and the treated teachers. Column (3) reports the difference and associated standard error for each item, adjusting for strata fixed effects. Column (4) adds pre-treatment teacher covariates.

Overall, I did not find large differences between the two groups. All of the teachers had a surprisingly poor understanding of the (even very basic) knowledge about college admissions policies. Treated teachers were 12.8 percentage points ($p<0.05$) more likely to report that they understood the Deferred Acceptance mechanism, while only 14.6 percent of teachers in the control group understood it. However, treated teachers were 13.3 percentage

[62] Table 2.21 presents the null effects on the itemized college match outcomes. Additionally, applying the "pay for percentile" method used in our performance pay contract, we estimate teacher-level "value-added" in students' college admissions outcomes, conditional on a student's CEE score ranking. Figure 2.11 shows that there are no large differences between the treated and the control teachers.

points (p<0.1) less likely to correctly answer the number of majors a student could apply to within each college. Overall, only 5 percent of control teachers correctly answered the three questions; but with 10 percentage points more among the treated teachers. About half of the teachers had performance pay based on student performance in the College Entrance Exam. However, there is no difference between the two groups largely due to our stratified (within school) randomization.

Figure 2.5 summarizes the descriptive results, showing suggestive evidence that teachers failed to provide effective instruction and advising to their students during the process of college choice and application. Most of the teachers did think that college application was important for students, but only fewer than 70 percent were confident in advising students with their college applications. More than half of the schools provided some form of whole-school advising activities, half of which were one-on-one advising or consultancy provided by these teachers. However, teachers were not knowledgeable about performing this task, even though they were all experienced with exam preparation. Interestingly, although students reported that high school teachers were their (most) important information sources, and teachers did provide advising and guidance, only about 20 percent of teachers found that they had an impact on students' actual college applications.

I supplemented the pay-for-performance contract with online training sessions. For teachers who did not take part in the online training sessions, they were provided links to the videos and other materials, including the guidebook. However, teachers may lack incentives to learn from our online training sessions and other supporting materials, resulting in low-take up of the training opportunities.[63] In evaluating the Evidence Based Literacy Instruction in elementary schools, Jacob (2017) finds null effects of teacher training on student achievement. He notes that providing teachers with a set of strategies and techniques, as I did in the college-going intervention, requires teachers to not only fully understand the techniques, but also to best use them. Survey data show that teachers reported a poor understanding of

[63]A future exploration is to examine how incentives affect their participation in training sessions or how different forms of training sessions would work.

some basic college application knowledge. It is then not surprising that they may not have done well in advising some more strategic and sophisticated decision-making , particularly the data analysis tasks in the targeting strategies.

Consistently, Table 2.22 shows that teacher incentives did not substantially affect students' college application behaviors. The results provide suggestive (statistically insignificant) evidence that students in the treated teachers' classes were more likely to follow the general nudge, and apply to colleges with higher tuition. These general guidelines could be shared during low-touch conversations. However, those students were no more likely to use the targeting strategies that required intensive data analytics. This mixed impact on students' college choices and applications may result in improvement on college access, but may reduce college match. Assistance from teachers may reduce a student's intrinsic motivation in making a thoughtful application. One supporting result is that the ITT effect of teacher incentives on student application time is -1.012 (p=0.077), suggesting that students in treated teachers' classes spent on average one hour less than students in the control group.

The ineffectiveness of the incentive policy may also be because we introduced a new performance measure without strong accountability pressure and close monitoring. Dee and Wyckoff (2015) show that the incentive plan works for teachers under career dismissal threat. In this chapter, the new performance measure depends upon productivity beyond those teachers' ability and skills in test preparation. The final performance ranking is uncertain. Teachers may be so ignorant of the education production function that they just maintain the status quo and hope for the best (Fryer, 2013). Therefore, there is a statistically zero difference between the treated and control teachers.

Supplemental evidence from 2018 survey data. For a few points discussed above, I did not have direct evidence. In late May 2018 I conducted a teacher survey with all the 20 classroom head teachers of the 2018 graduate cohort in a randomly chosen high school. I asked in-depth questions about college choice advising in that high school. For the common questions with the 2017 survey, I found very similar results that high school teachers thought college application is important to students, but did not prove to have a good understanding of college admissions. Half of the teachers agreed that their suggestions were very important to students, but only 5% of students (N=1,190, in the student survey) reported that teachers' suggestions and guidance were very important in affecting their college choices.

Around half of the teachers provided class workshops or one-on-one advising to students. They reported that one limitation was lack of teachers experienced in college-going advising. Regarding school improvement, 95% of teachers thought that high schools should improve their teaching quality, and 65% of teachers thought that high schools should provide better college-going advising services. In contrast, fewer than 40% of teachers thought that high schools should have more sports activities, or upgrade their facilities. About half of the teachers agreed that high school principals and teachers had better knowledge than students and parents regarding college choices and applications. On the other hand, half of the teachers thought that either admissions policies or parents were more important than schools in improving college choices.

Heterogeneous treatment effects. The characteristics of advisers may affect the effectiveness of personalized advising programs. I directly tested for this possibility in Table 2.9. I did not find statistically significant heterogeneous effects by teaching experience, earnings, and belief about the importance of college choice and application. The point estimates suggest that teachers who were younger, had higher monthly salaries, and thought college application is important, outperformed other teachers. The last pair of columns show that teachers who reported to have *ex post* impacted their students' college choices actually harmed student college access, because students were more likely to retake the CEE in the next year. Table 2.23 presents no evidence of statistically significant heterogeneity on dimensions of student characteristics. In Figure 2.6, I tested for equality of treatment effects for quintiles of the CEE score distribution. Results show that teacher incentives did not benefit students at any part of the distribution. There is suggestive evidence that low-achieving students benefited more than high-achieving students, but the estimates were not precise enough.

2.7 Discussion and Policy Implications

In this chapter, I search for, design, and examine effective behavioral interventions to improve low-income students' college access and match at scale. The findings have significant implications for policy, practice, and research. The poverty gap in college access and match has persisted for decades all over the world. We are far from finding the final policy solutions for millions of low-income students. More research and policy efforts are needed.

This chapter focuses on the academic match between students and colleges because it provides a reference point for the desired outcome: We would like all qualified students to go to college and to go to the one that best fits their academic abilities and preferences. On average, academic undermatch negatively impacts students' college and longer-term outcomes. Nevertheless, we should note that a student may have other needs or preferences, and her undermatched college may be a good fit for her (Smith *et al.*, 2013; Bond *et al.*, 2018). Providing information and knowledge that shapes students' motivations and preferences is still crucial to help students make their way to their "fit" colleges.[64] Using the general knowledge of decision-making (specifically in college choice) and more college and major information, students could research, find, and rank their match or fit colleges.

District and school leaders are under pressure to improve college-going outcomes, especially among disadvantaged students. With limited school financial and human resources, improving student decision-making (and many other school decisions) through behavioral interventions such as information, knowledge, and personalized assistance seems to be a promising policy tool to reach these goals. As a result of behaviorally-informed designs, unique access to data, and returns to scale, the costs of the individualized advising are relatively low. The per student cost of the expert human instruction in 2016 is about $30,

[64]For example, Muslim students in our experimental sample have incorrect beliefs that they would face dietary restrictions if they attended out-of-province colleges.

and that of the machine learning-assisted advising in 2017 is about \$20. The personalized advising cost could be further reduced by using artificial intelligence and other new technologies (Page and Gehlbach, 2017). Additionally, as there is a fixed development cost of the machine learning system, the average cost will largely drop if the number of users increase at a minimal marginal cost (in this experimental case, as a result of increased take-up). [65]

The average per student cost of the other design - teacher incentives - is also low, ranging from to \$3 to \$10 (if all the treated teachers receive the performance pay); however, its impact in 2017 was zero. The null results suggest that, while school personnel policies such as pay-for-performance could potentially incentivize teachers, the incentive design without complementarities such as school inputs and organizational policies (e.g., accountability-based mandatory training) likely undermines its effectiveness.[66]

Many behavioral interventions that have been designed and tested by researchers are in the family of public services. In both centralized and decentralized college admissions systems, providing a centralized data system would greatly reduce search costs for students. However, in many cases, access to information and knowledge may not be sufficient for students to make informed choices; instead, the use of information and knowledge matters. Even with such a centralized data system, as noted by Hayek (1945), the ultimate decisions must be left to the people who are familiar with the changes in the particular circumstances. In the case of college choice, the personalized decisions should not be determined merely by a central board, but by each student with their own beliefs and preferences.

Data- and technology-based methods are increasingly useful for policy prediction problems. These new methods, used jointly with expertise and data, have wide application prospects in increasing school effectiveness along the educational pipeline before and beyond

[65] The cost of the ineffective "business as usual" advising is low (about \$0.5 per student) as we only need to pay for the working hours of the counselors. For scale-up, the province-centralized college application system could develop such a chatbot or an online advising system to serve all applicants.

[66] For example, we have implemented a pay-for-performance intervention for principals in Yunan Province in 2017, using both monetary and non-monetary (career promotion) incentives. Though not examined, increasing parental engagement would also potentially improve student outcomes (Bergman, 2015; Castleman and Page, 2017).

the transition from K-12 schools to college. Future work is needed to test how to combine "human instruction" and "machine learning" to make the best possible policy decisions in schools and colleges. Furthermore, while the interventions are not designed to recruit students to a particular college, college administrators could improve their recruitment efforts and performance using similar behavioral interventions (Castleman *et al.*, 2015a; Dynarski *et al.*, 2018; Miller and Skimmyhorn, 2018).

One limitation of this chapter is the low take-up rate, which limits us from analyzing the intervention effects and effect heterogeneities on a large scale. Low take-up rates are a common problem in social programs, especially those implemented by researchers (Currie, 2004). It is well documented that external experts may have smaller impacts than their trusted ones (on a regular basis) on program participation and eventual outcomes (Bertrand *et al.*, 2004; Chetty, 2015). For a policy intervention on a large scale, a key issue is to raise take-up or willingness to pay, in order to increase the program effectiveness. It may be of particular policy and research interest to increase the take-up among those who need the policy most. In the case of college-going interventions, this represents a common issue facing researchers in both the U.S. decentralized systems and the Chinese centralized systems. Students and parents often suspect that, even if (or in particular) all the services are free, researchers must have other purposes such as selling college advice or collecting individual data. This fact should motivate researchers to consider a closer collaboration with district and school leaders in order to give the intervention projects "a strong, public presence" (Hoxby and Turner, 2013).

In transforming the RCT results in this chapter into a scaled policy intervention, we need to consider the replicability of the intervention design and the general equilibrium effects that may arise at scale. First, two years in a row, Chapter I and Chapter II find evidence that the intervention impact was not a fluke. If students fully follow the structured individualized advising and the machine learning-based predictions, we would expect the average treatment effect on college choice behaviors in general equilibrium to be the same as the average treatment effect on the treated as identified in this chapter.[67] Second, the main concern about the general equilibrium effects is that, as college admission is a zero-sum game that students compete for the scarce spots in each college, the scaled-up effect on college access and match might be smaller than the RCT effect. However, the redistributional effect is theoretically ambiguous: there could be positive spillover effects from coordination, or negative crowd-out effects from competition.[68] A lot of more data and work is needed to identify these general equilibrium effects and to improve the take-up, targeting, and effectiveness of the college-going interventions.

2.8 Conclusion

I have asked how to effectively improve college access and match at scale in centralized college admissions systems. In 2017, I conducted a set of field experiments among the universe of high school graduates in one of the poorest provinces in China. Using unique administrative data on student college application and admission, this chapter identifies the causal effects of various scalable behavioral policy interventions on students' college choice behaviors and admissions outcomes.

This chapter has two clear findings. First, application assistance regarding knowledge and strategies in college choice is effective in improving college access and match among some low-income students. Learning and using the knowledge of how to make optimal college choices is the key element that shapes a student's college application behaviors without changing their preferences for colleges and majors.

Second, I have tested two approaches that would potentially scale up the labor-intensive

[67] In non-experimental settings, we have provided online advising to hundreds of thousands of high school graduates in different Chinese provinces and have received consistently positive feedback. The fundamental elements of the six-step structured advising works in different contexts (e.g., variations in the admissions mechanisms). Although students from economically advantaged families or regions have better information and resources, they often have misunderstanding about the admissions mechanisms and strategies as well that the advising program is likely to produce similar impacts.

[68] As many colleges accentually admit fewer students than their designated quotas, improvement in students' college choices would potentially benefit both colleges and students through getting more students into college.

individualized advising program: (1) increasing advising efficiency using big data-based methods, and (2) increasing the supply of advisors using teacher performance pay policy. Machine learning algorithms suggest a potential policy lever to scale up the college-going advising programs and other in-kind data-based behavioral interventions in both K-12 and higher education. However, without complementary school organizational policies and sufficient teacher training, pay-for-performance contracts alone do not meaningfully incentivize teachers to provide effective college application assistance. Teachers remain as potential labor resources to assist students' college choices at scale, but how to best utilize these idled human resources needs further investigation.

More generally, the results suggest that behaviorally motivated interventions and policies - in many cases simple and inexpensive - have a high potential to improve personalized predictions and decision-making in education and other aspects of life, particularly for low-income and underrepresented minority students. Many such interventions are public goods. Policymakers and schools/colleges could scale up the practices through improving existing K-12 and higher education systems, while also bringing in the power of behavioral interventions for teachers and students, big data methods, and evidence-based policy designs.

2.9 Main Figures

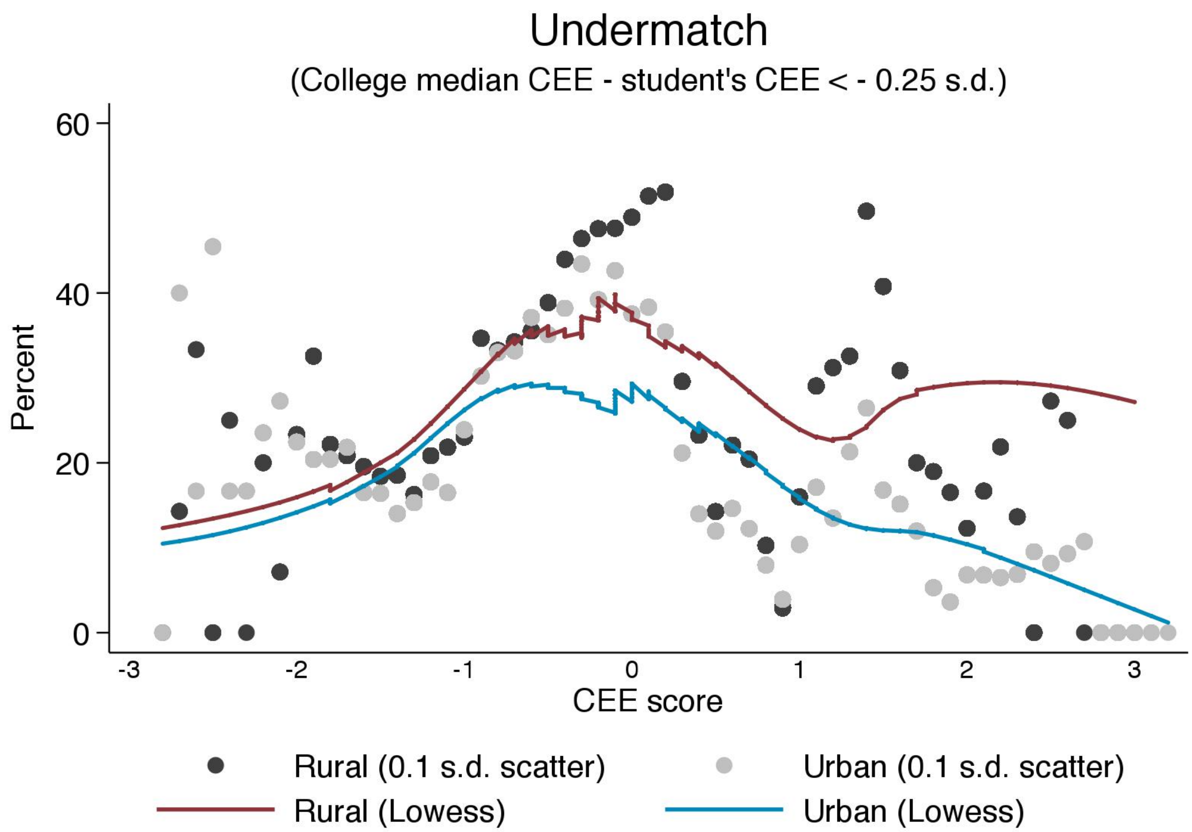
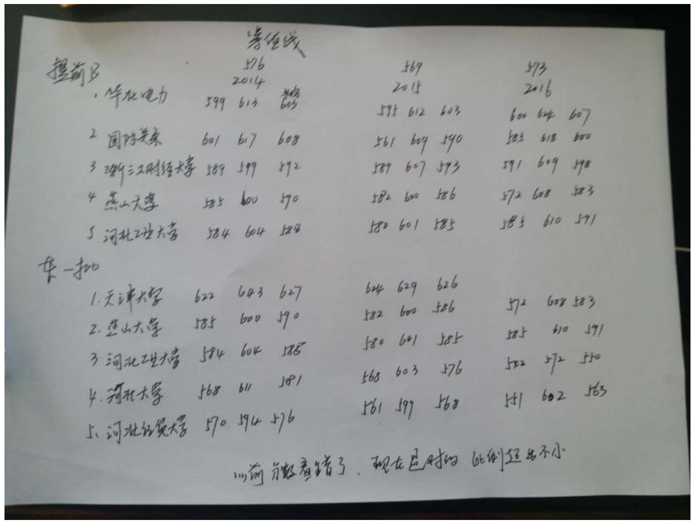


Figure 2.1. The poverty gap in academic match

Notes: This figure plots locally weighted regression lines of academic undermatch rates between rural (proxy of poor) and urban students on the x-axis of students' CEE score (standardized). Each dot represents the average undermatch rate for a 0.1 s.d. bin. Undermatch is defined as a student being admitted by a college with median CEE score 0.25 s.d. lower than her own CEE score or not being admitted by any college. The sample includes the universe of the untreated sample of the 2017 cohort of high school graduates in Ningxia, China. Average undermatch rate among rural students is 31% and that among urban students is 23%.



(a) Manual score equating

```
Sunday June 25 21:02:26 2017   Page 1
```

Stata(R) Statistics/Data Analysis

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

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

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

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

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

(b) College short list in 2017

Figure 2.2. How does *Stata* reduce human labor?

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

(a) A low predicted probablity

(b) A high predicted probability

Figure 2.3. How does machine learning work in 2017?

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

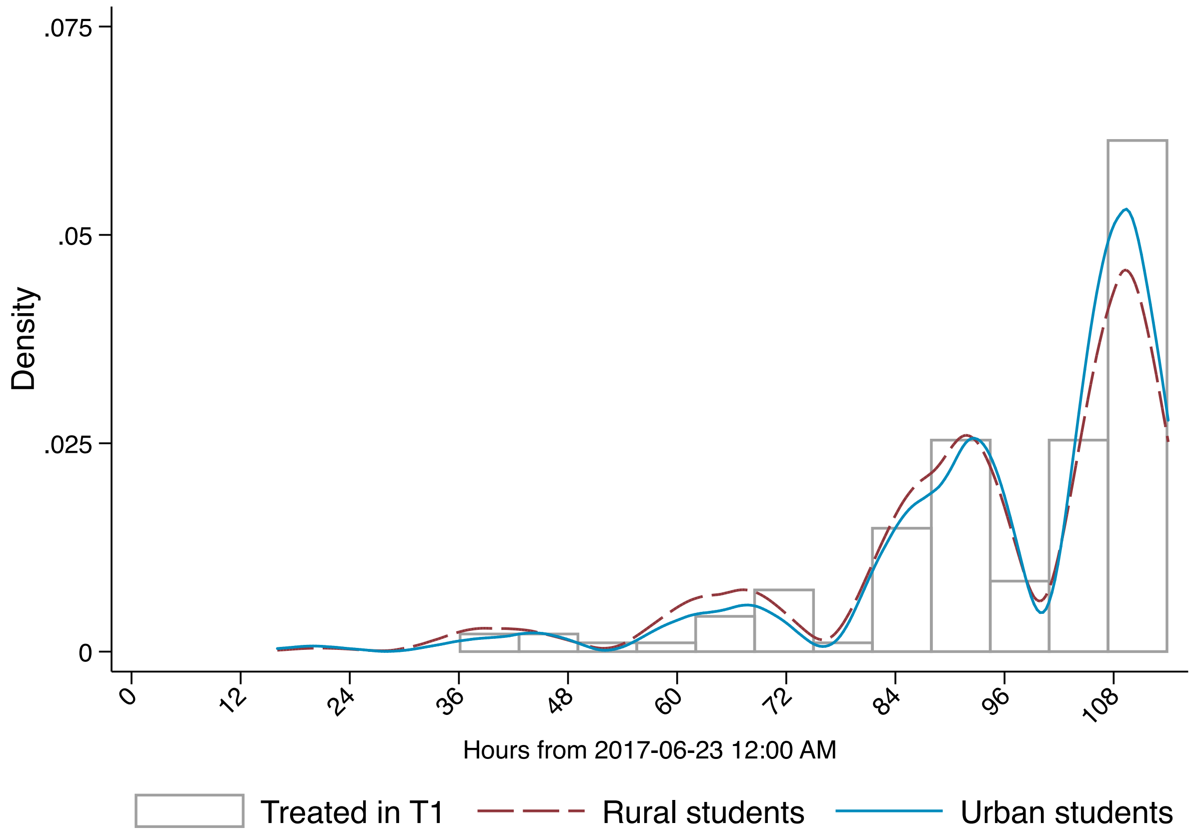


Figure 2.4. Distribution of college application submission time

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

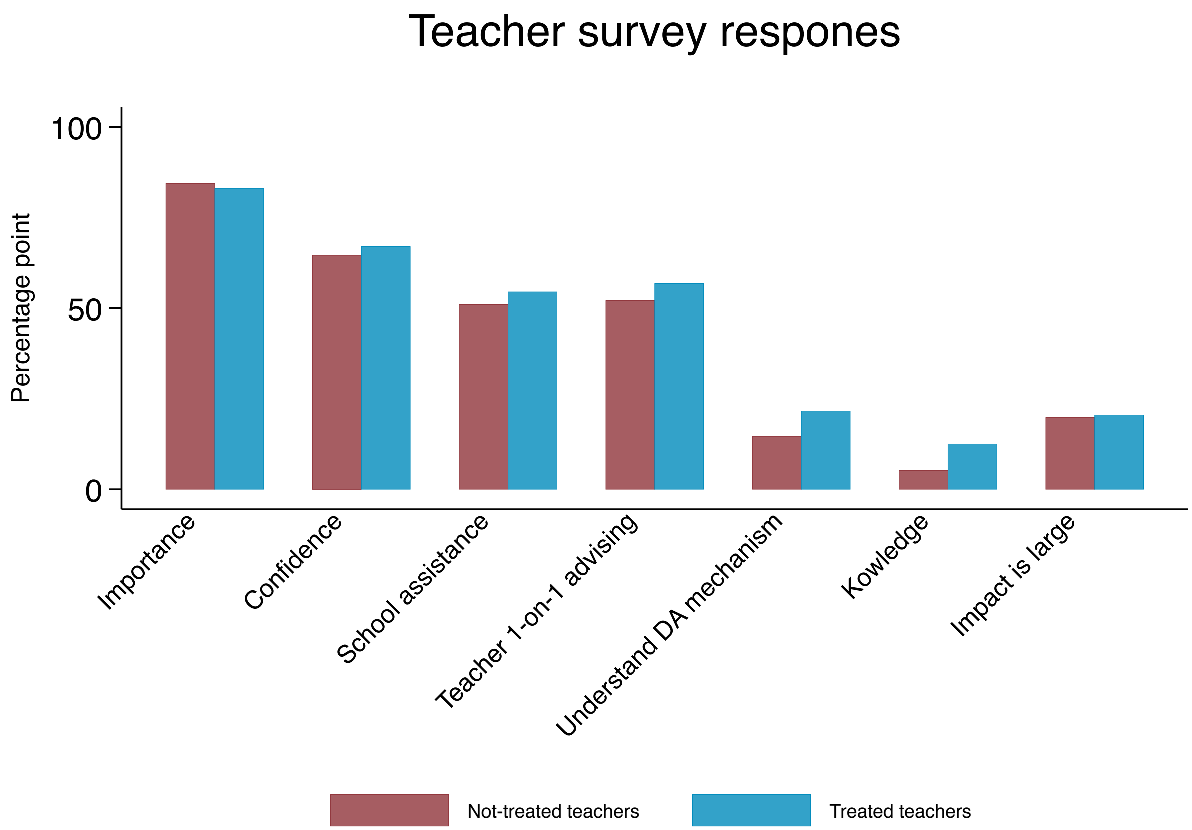
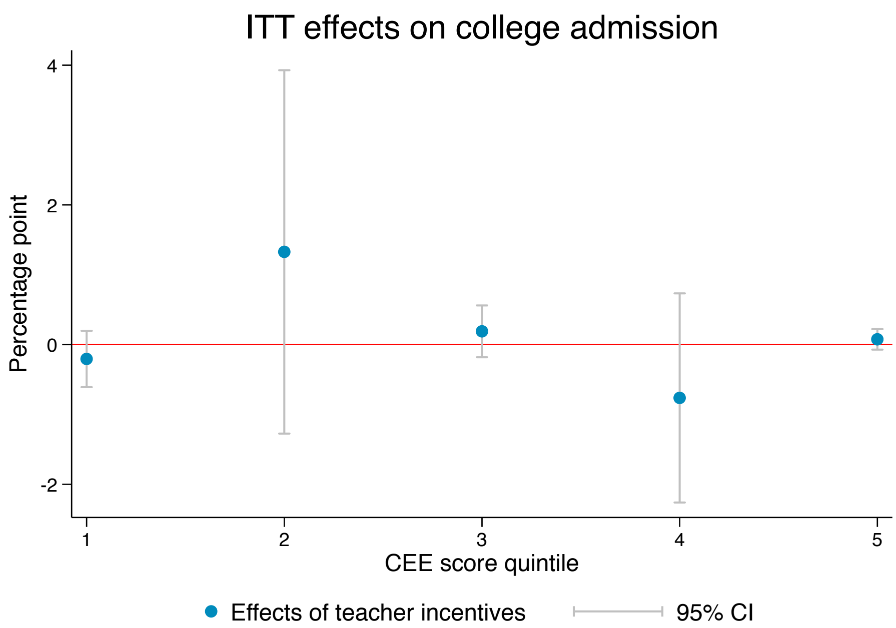
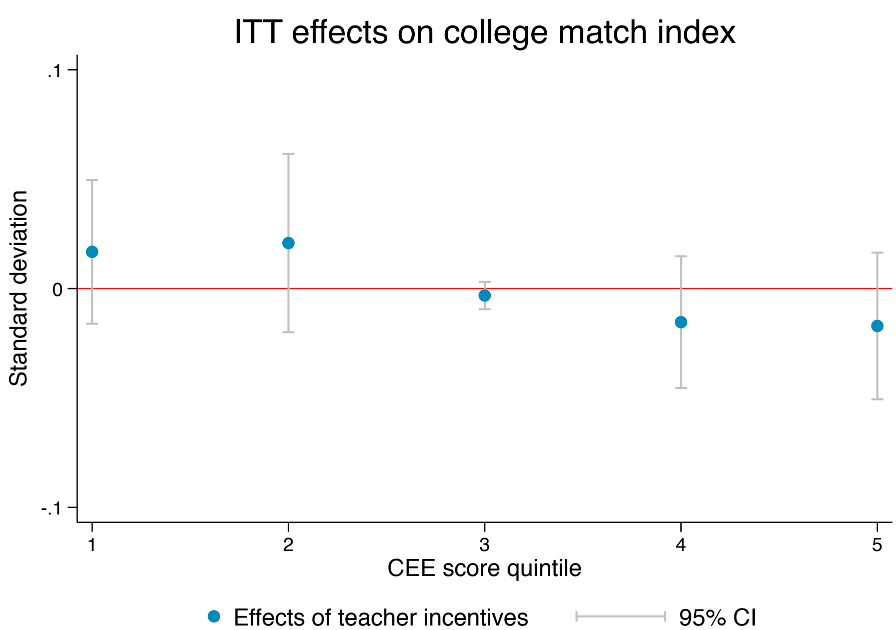


Figure 2.5. Teacher survey responses

Notes: This figure shows the sample average of each dichotomous answer to the post-treatment survey questions for the control group and the treatment group, respectively. **Importance:** "College application is important." **Confidence:** "I am confident in helping students with college application." **School assistance:** "Our high school provided whole-school college application assistance (in any forms)." **Teacher 1-on-1 advising:** "I provided 1-on-1 advising to students in my class." **Understand DA mechanism:** "I fully understand the Deferred Acceptance mechanism." **Knowledge:** Understand DA mechanism + correct answers about the number of colleges and majors that a student could apply to. **Impact is large:** "I believe I had large impacts on students' college applications in my class."



(a) Effects on college admissions



(b) Effects on college match (standardized index)

Figure 2.6. Heterogeneity in the ITT effects by student CEE score distribution

Notes: This figure plots heterogeneous ITT effects of teacher incentives on college access and match by student CEE score quintile distribution. Results are from separate regressions of Model (2.5) using subsamples. Average treatment effects for admission and match index are 0.004 (0.008) and -0.005 (0.017) (columns 1 and 5 in Table 2.7).

2.10 Main Tables

Table 2.1. Experimental design: Interventions in 2017

	A. Student interventions in 2017				
Question	Does machine learning improve personalized advising efficiency?				
Sample	All 54,055 students in 61 public high schools				
Randomization	Student level within 887 strata				
Groups	Intervention	Communication	Randomization unit (students)	Analysis unit (students)	Take-up
Control	No		43,038	43,038	
Treatment 1	ML-assisted advising	Official text message	5,647	5,647	3.6%
Treatment 2	"Business as usual" advising	Official text message	5,370	5,370	2.4%

	B. Teacher intervention in 2017				
Question	Does pay-for-performance incentivize teachers to provide effective college application assistance?				
Sample	9,354 students in 184 head teachers' classes				
Randomization	Teacher level within 49 strata				
Groups	Intervention	Communication	Randomization unit (teachers)	Analysis unit (students)	Take-up
Control	No		96	4,865	
Treatment 1	Teacher incentives	Official contract letter	88	4,489	100%

Notes: This table shows the experimental design of the *Bright Future of China Project* in Ningxia in 2017. Randomization strata of student interventions are by school, track, gender, race, rural hukou, county of residence, and achievement (classifying high-achieving students using low-stakes graduation test scores). Randomization strata of teacher intervention are by school, track, and elite class categories. Both the text messages and contract letters were sent through the official channels of Ningxia Department of Education and the research center at Peking University.

Table 2.2. College choices and the poverty gap in college match

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

Notes: This table reports the OLS regression (extension of Model (2.1)) results for the partial correlations between college application behaviors and the college match index (standardized), using data from those who submitted college applications in the control group in 2017. Application behaviors are constructed using the full applications data, as described in Appendix subsection 2.13.1. Columns (1) and (2) present the mean values of each college behavior index for rural and urban students. Column (3) shows the rural-urban gap in college admissions using the full untreated sample (same as in column (3) of Panel C in Table 2.10). The next three columns add the strategy and preference measures (principal component factor indices) stepwise. Columns (7)-(10) control for high school fixed effects. All regressions include a student's CEE score and other demographic covariates. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.3. First stage: Take-up of individualized advising programs

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

Notes: This table reports the OLS regression (Model (2.3)) results of the take-up of the individualized advising interventions in 2017. Over 1,800 users (students or parents; about 16% of the treatment group size) added us as friends in the online message App (WeChat), but many of them refused to provide their exam ID and school ID for verification. Students in both groups received the same text message (the only exception is the contact information). Take-up rates among high-achieving students were 4.8% and 4.1% for the two treatment groups. All regressions control for strata fixed effects. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.4. ITT and TOT effects on college access and match of the individualized advising programs

	Admission (=1)		Application (=1)		Enrollment in 2017 (=1)		Repeating in 2018 (=1)	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1)		(2)		(3)		(4)	
Control mean	0.837		0.900		0.765		0.209	
Control sd	[0.369]		[0.301]		[0.424]		[0.407]	
T1 (machine learning)	0.009*	0.242*	0.005	0.128	0.010*	0.269*	-0.009*	-0.233*
	(0.005)	(0.130)	(0.004)	(0.105)	(0.005)	(0.149)	(0.005)	(0.138)
T2 (business as usual)	-0.006	-0.254	-0.004	-0.143	-0.010**	-0.390**	0.008	0.338
	(0.004)	(0.173)	(0.004)	(0.147)	(0.005)	(0.191)	(0.005)	(0.220)
$\Pr(\beta[T1] = \beta[T2]=0)$	0.066	0.057	0.305	0.284	0.037	0.032	0.074	0.066
$\Pr(\beta[T1] = \beta[T2])$	0.021	0.021	0.127	0.127	0.011	0.009	0.024	0.025
	Index (s.d.)		College quality (s.d.)		Index drop non-admitted (s.d.)		Undermatch (=1)	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(5)		(6)		(7)		(8)	
Control mean	-0.069		-0.170		-0.085		0.280	
Control sd	[0.975]		[1.202]		[0.980]		[0.449]	
T1 (machine learning)	0.022**	0.598**	0.028**	0.770**	0.010	0.262	-0.011*	-0.288*
	(0.009)	(0.255)	(0.013)	(0.360)	(0.007)	(0.173)	(0.006)	(0.166)
T2 (business as usual)	-0.007	-0.265	-0.009	-0.337	0.002	0.081	0.007	0.285
	(0.008)	(0.330)	(0.012)	(0.483)	(0.005)	(0.189)	(0.006)	(0.226)
$\Pr(\beta[T1] = \beta[T2]=0)$	0.042	0.035	0.077	0.066	0.311	0.303	0.056	0.048
$\Pr(\beta[T1] = \beta[T2])$	0.018	0.027	0.035	0.053	0.336	0.466	0.017	0.019

Notes: This table reports the OLS regression results of the ITT effects (Model (2.2)) and TOT effects (Model (2.4)) of the individualized advising interventions in 2017 on a family of college access and match outcomes. The sample includes the universe of public high school graduates in Ningxia in 2017 (N=54,055). Column (7) only includes students who were admitted to college (N=45,482). **Machine learning** intervention used the predicted probabilities of admissions to each college-major-order for each student and other automatic data analyses to assist the conventional online personalized advising. **Business as usual** intervention provided low-touch, brief college application guidelines and tips to students, which is used as a placebo test that mimics a low-price for-profit consulting service available to students if they are willing to pay for it. **Index** denotes a principal-component index of college quality using information from five measures (median, mean, and minimum admissions scores; national college ranking scores and percentiles). **College quality** is the national college ranking score (standardized) using college admissions data from 1996-2017 and administrative data on institutional resources for every college in China. The other outcomes are dichotomous variables. All regressions control for student-level covariates and strata fixed effects. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.5. ITT and TOT effects of individualized advising programs: College application behaviors

	Apply to colleges in ML list		College application Index		Strategy Targeting		Strategy General nudge	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1)		(2)		(3)		(4)	
Control mean	0.305		-0.030		-0.028		-0.029	
Control sd	[0.460]		[1.001]		[0.994]		[1.003]	
T1 (machine learning)	0.011**	0.306**	0.022*	0.599*	0.033***	0.894***	0.030**	0.804**
	(0.006)	(0.150)	(0.012)	(0.332)	(0.011)	(0.305)	(0.012)	(0.338)
T2 (business as usual)	-0.003	-0.114	-0.018	-0.728*	-0.026**	-1.063**	-0.012	-0.489
	(0.006)	(0.225)	(0.011)	(0.439)	(0.013)	(0.491)	(0.012)	(0.482)
$\Pr(\beta[T1] = \beta[T2]{=}0)$	0.136	0.119	0.126	0.101	0.007	0.005	0.058	0.049
$\Pr(\beta[T1] = \beta[T2])$	0.099	0.139	0.043	0.035	0.002	0.002	0.035	0.042

	Strategy Special programs		Preference Tuition & quota		Preference Location		Preference Major	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(5)		(6)		(7)		(8)	
Control mean	-0.030		0.030		-0.058		0.003	
Control sd	[0.978]		[1.012]		[0.977]		[0.999]	
T1 (machine learning)	-0.005	-0.137	0.018	0.484	-0.005	-0.137	-0.002	-0.058
	(0.017)	(0.456)	(0.014)	(0.375)	(0.011)	(0.294)	(0.016)	(0.425)
T2 (business as usual)	-0.000	-0.010	-0.001	-0.014	-0.003	-0.113	-0.009	-0.354
	(0.012)	(0.499)	(0.011)	(0.448)	(0.010)	(0.395)	(0.017)	(0.672)
$\Pr(\beta[T1] = \beta[T2]{=}0)$	0.956	0.955	0.464	0.434	0.878	0.874	0.864	0.861
$\Pr(\beta[T1] = \beta[T2])$	0.801	0.831	0.322	0.390	0.879	0.958	0.774	0.714

Notes: This table reports the OLS regression results of the ITT effects (Model (2.2)) and TOT effects (Model (2.4)) of the individualized advising interventions in 2017 on a family of college application behaviors. The sample includes the universe of public high school graduates in Ningxia in 2017 (N=54,055). **Machine learning** intervention used the predicted probabilities of admissions to each college-major-order for each student and other automatic data analyses to assist the conventional online personalized advising. **Business as usual** intervention provided low-touch, brief college application guidelines and tips to students, which is used as a placebo test that mimics a low-price for-profit consulting service available to students if they are willing to pay for it. See the text for more descriptions of the outcome variables. All regressions control for student-level covariates and strata fixed effects. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.6. Heterogeneity in the ITT effects on application time use

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

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

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

Table 2.7. ITT effects of teacher incentives: College access and match outcomes

	Access outcomes				Match outcomes			
	Application (=1) (1)	Admission (=1) (2)	Enrollment (s.d.) (3)	Repeating (=1) (4)	Index (s.d.) (5)	Quality (s.d.) (6)	Index* (s.d.) (7)	Undermatch (=1) (8)
Control mean	0.902	0.835	0.738	0.238	-0.030	-0.118	-0.027	0.279
Control s.d.	[0.297]	[0.371]	[0.440]	[0.426]	[1.011]	[1.250]	[1.019]	[0.449]
A. Without teacher covariates								
Teacher incentives	0.004	0.002	0.002	0.001	-0.005	-0.008	-0.013	0.015
	(0.008)	(0.011)	(0.011)	(0.011)	(0.017)	(0.025)	(0.010)	(0.012)
	¡0.009¿	¡0.012¿	¡0.013¿	¡0.014¿	¡0.018¿	¡0.025¿	¡0.010¿	¡0.012¿
B. With teacher covariates								
Teacher incentives	0.007	0.004	0.006	-0.005	-0.001	-0.001	-0.011	0.015
	(0.008)	(0.009)	(0.009)	(0.010)	(0.016)	(0.024)	(0.010)	(0.010)
	¡0.009¿	¡0.011¿	¡0.013¿	¡0.014¿	¡0.018¿	¡0.024¿	¡0.009¿	¡0.011¿
N	9,354	9,354	9,354	9,354	9,354	9,354	9,354	9,354

Notes: This table reports the OLS regression (Model (2.5)) results of the ITT effects of the pay-for-performance incentives for teachers in 2017 on a family of college access and match outcomes. All regressions control for student-level covariates and strata fixed effects. Teacher covariates include both aggregated student covariates in each teacher's home class and teacher characteristics in the pre-treatment survey data, as described in Table 2.13. Standard errors in parentheses (angle brackets) are clustered at high school (teacher) level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.8. Measuring teachers' belief, knowledge and action in college application advising

	Mean		Treatment effect	
	Control	Treatment	no control	with control
	(1)	(2)	(3)	(4)
College application is important (=1)	0.844	0.830	-0.011	-0.012
	[0.365]	[0.378]	(0.056)	(0.057)
Confident in advising college application (=1)	0.646	0.670	0.047	0.034
	[0.481]	[0.473]	(0.074)	(0.074)
Understand DA mechanism (=1)	0.146	0.216	0.112*	0.128**
	[0.355]	[0.414]	(0.059)	(0.065)
Correct answer: # colleges (=1)	0.521	0.580	0.042	0.074
	[0.502]	[0.496]	(0.079)	(0.078)
Correct answer: # majors (=1)	0.396	0.250	-0.135*	-0.133*
	[0.492]	[0.435]	(0.072)	(0.072)
Kowledge (=1)	0.052	0.125	0.091*	0.101*
	[0.223]	[0.333]	(0.049)	(0.053)
School-level advising activities (=1)	0.510	0.545	0.013	-0.008
	[0.503]	[0.501]	(0.078)	(0.080)
Teacher 1-on-1 advising (=1)	0.521	0.568	0.061	0.043
	[0.502]	[0.498]	(0.076)	(0.076)
Teacher's impact is large (=1)	0.198	0.205	0.039	0.052
	[0.401]	[0.406]	(0.064)	(0.067)
Test score performance pay in salary (=1)	0.490	0.500	-0.008	0.005
	[0.503]	[0.503]	(0.081)	(0.080)
Missing responses	0.052	0.034	-0.012	-0.002
	[0.223]	[0.183]	(0.026)	(0.024)
Schools	23	25		
Teachers	96	88		

Notes: This table reports the mean differences in post-treatment teacher survey items. Column (1) and (2) report the mean and standard deviation of each question item for the control and treatment groups, respectively. Column (3) reports the treatment-control difference and associated robust standard error from an OLS regression controlling for strata fixed effects. Column (4) adds pre-treatment teacher covariates. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.9. Heterogeneity in the ITT effects of teacher incentives: By teacher characteristics

	Experience		Monthly salary		Application is important		Impact on students	
	¡10 years	¿=10 years	¡3000	¿=3000	No	Yes	No	Yes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
N (students)	3,533	5,821	4,263	5,091	5,338	4,016	7,485	1,869
Admission	0.005	0.000	-0.012	0.014	0.003	0.002	0.010	-0.028**
	(0.015)	(0.012)	(0.014)	(0.013)	(0.013)	(0.018)	(0.012)	(0.012)
College match index	0.004	-0.012	-0.027	0.013	-0.011	0.001	0.008	-0.055**
	(0.024)	(0.019)	(0.027)	(0.017)	(0.028)	(0.024)	(0.019)	(0.024)
College choice index	0.018	0.002	-0.001	0.016	0.029	-0.014	0.028	-0.069*
	(0.040)	(0.030)	(0.039)	(0.031)	(0.038)	(0.043)	(0.030)	(0.035)
Targeting strategy	0.011	-0.023	-0.002	-0.015	-0.008	-0.009	-0.001	-0.041
	(0.032)	(0.029)	(0.035)	(0.028)	(0.037)	(0.038)	(0.027)	(0.030)

Notes: This table reports the OLS regression (Model (2.5)) results of the heterogeneous ITT effects of the pay-for-performance incentives for teachers in 2017 on college match index. Measure of **Impact on students** is from the post-treatment survey. All regressions control for student-level covariates and strata fixed effects. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

2.11 Additional Figures

The Bright Future of China Project (Ningxia) Timeline

2015/10	Proposal to Ningxia Department of Education
2016/03	Approval of the proposal
2016/04	Data collection and randomization
2016/06	*College Entrance Exam*
2016/06	Interventions (during *college applications*): Guidebook, workshop, 1-on-1
2016/08	*College admissions*
2017/01	Data collection
2017/02	New proposal approval
2017/05	Teacher survey
2017/06	*College Entrance Exam*
2017/06	Interventions (during *college applications*): Teacher incentives, 1-on-1 using ML
2017/08	*College admissions*
2018/01	*Data collection*

Ongoing...

Figure 2.7. Project timeline

Notes: This figure plots the general timeline of the Bright Future of China Project (Ningxia). Interventions in both 2016 and 2017 were primarily implemented during the short college application periods. Individualized advising was also provided in early August when students applied to non-selective colleges.

Figure 2.8. Location of Ningxia

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

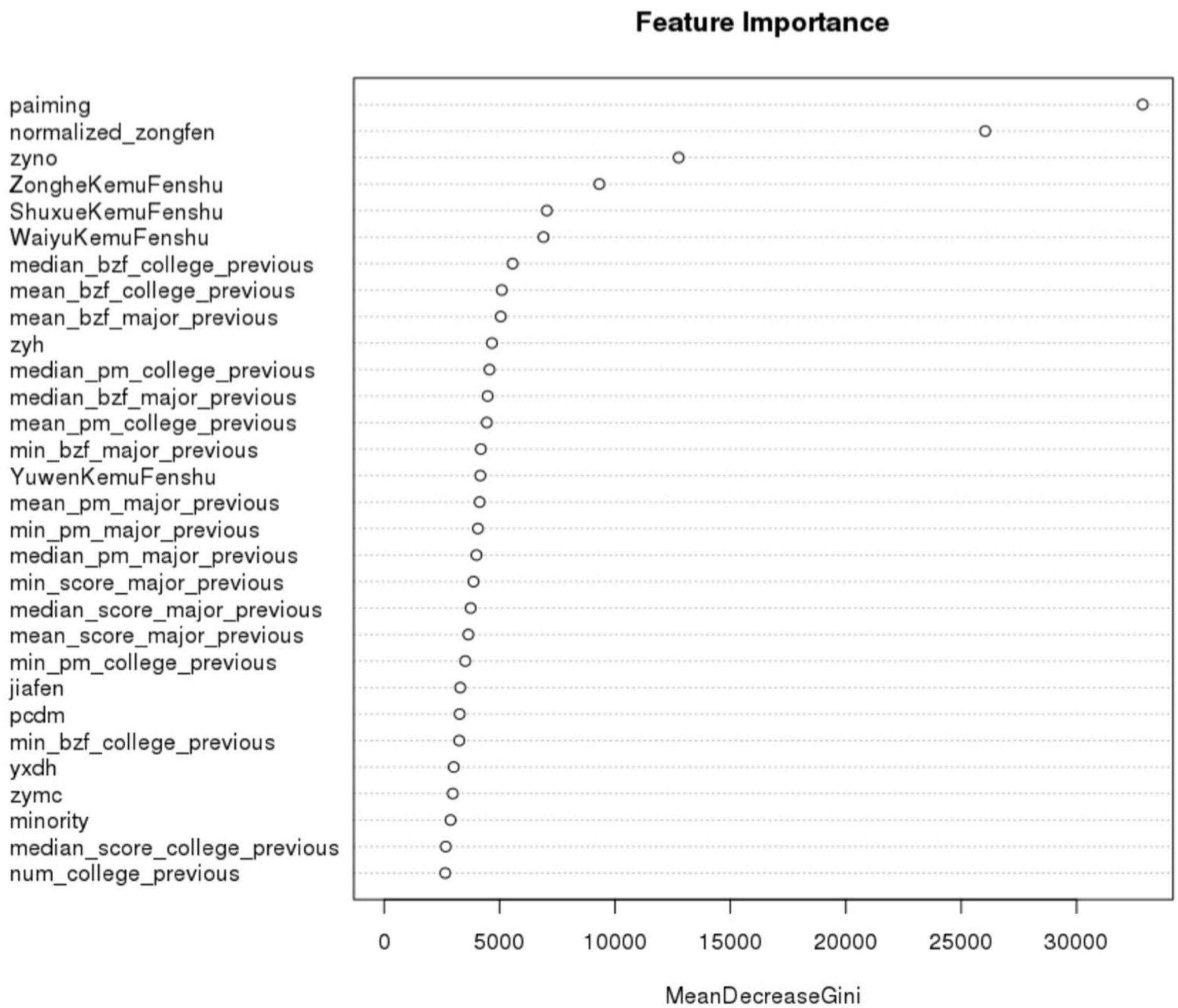


Figure 2.9. Relative importance of features in the random forest predictions

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

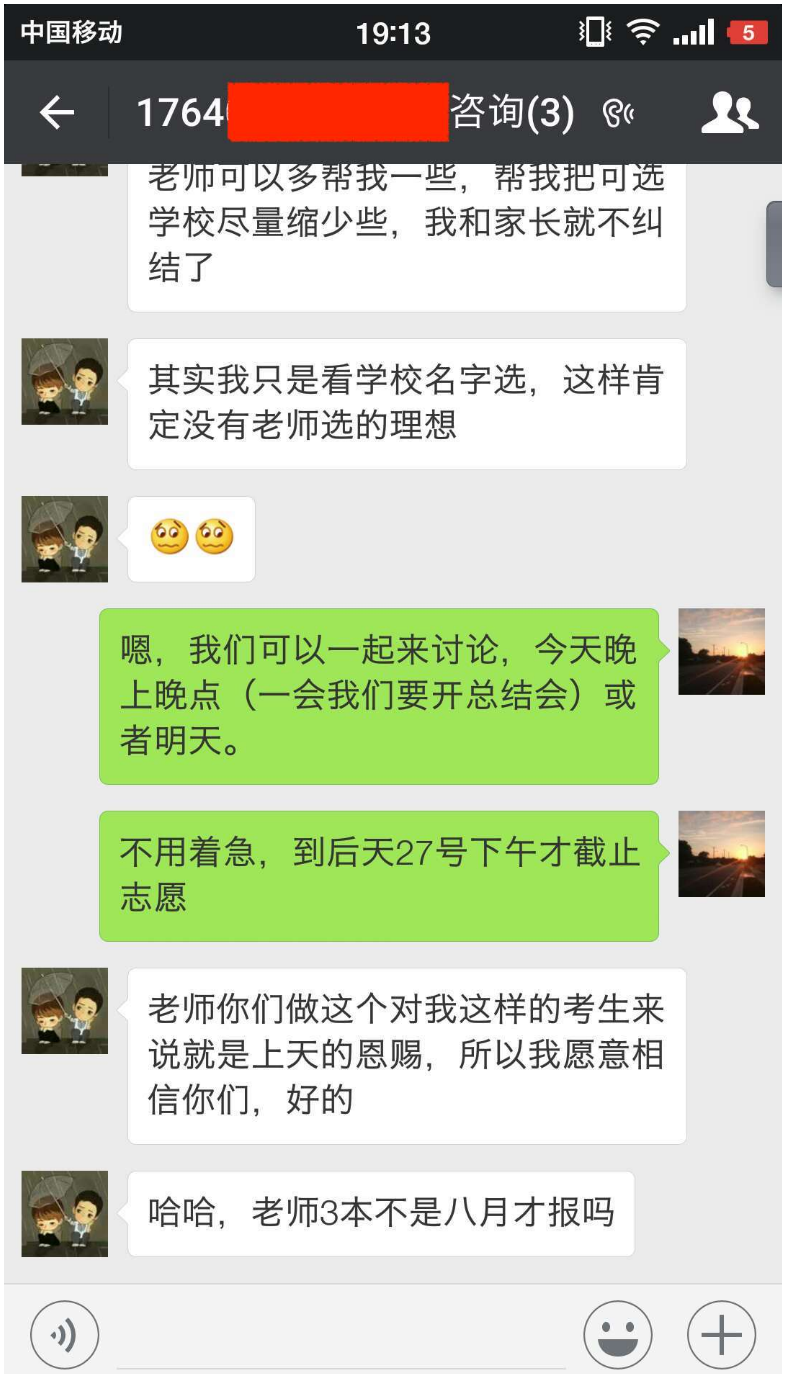


Figure 2.10. Example of the online individualized advising in 2017

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

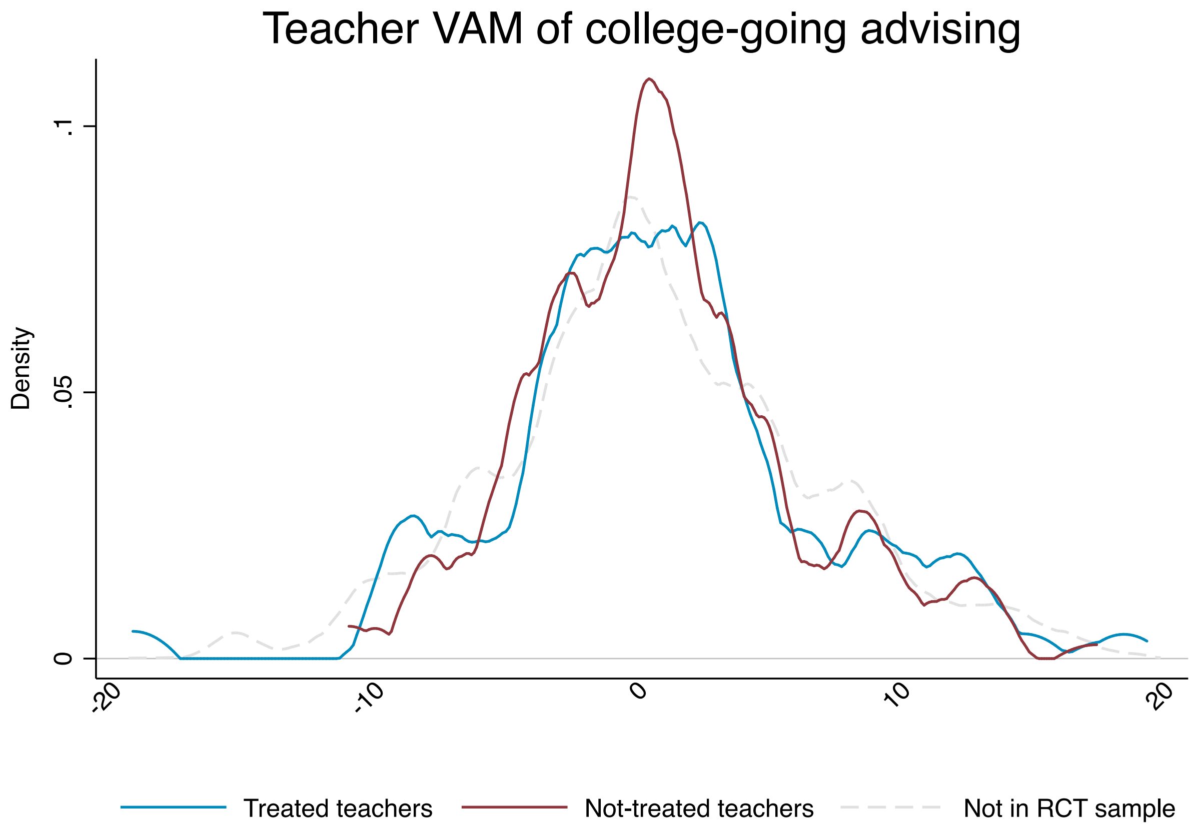


Figure 2.11. Distribution of teachers' value-added of college-going advising

Notes: This figure shows the density distribution of individual teacher's "value-added" in college admissions outcomes, holding students' CEE score constant, separately for the treated teachers, control teachers, and the teachers not in the experimental sample. The "value-added" is estimated by averaging the difference between college median score rank percentile and CEE score rank percentile for each student.

2.12 Additional Tables

Table 2.10. The poverty gap in college match

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

Notes: This table reports the OLS regression (Model (2.1)) results of the rural-urban gap in college match outcomes (as being summarized in the single index), using the control group sample in 2017. Columns (3) and (4) exclude students who were not admitted to a college. Columns (4) controls for high school fixed effects. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.11. Balance checks for student interventions

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

Notes: This table reports the balance checks results using student-level data in 2017. There were 836 treated students in treated teachers' classes. Joint F test results are from regressions in Table 2.12. Strata fixed effects are included. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.12. Balance checks for student interventions: Prediction of treatment status

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

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

Table 2.13. Balance checks for teacher interventions: Teacher-level covariates

	Control (1)	Treatment (2)	T-C diff (3)
Experience (years)	13.753	12.863	-1.103
	[7.405]	[7.331]	(1.053)
Missing experience	0.115	0.091	-0.023
	[0.320]	[0.289]	(0.044)
Monthly salary	3,726.246	3,408.937	-39.189
	[987.492]	[768.708]	(255.191)
Missing "salary"	0.323	0.284	-0.028
	[0.470]	[0.454]	(0.062)
College choice is important	0.400	0.475	0.110
	[0.493]	[0.503]	(0.078)
Missing "important"	0.115	0.091	-0.023
	[0.320]	[0.289]	(0.044)
Advanced certificate	0.529	0.525	-0.021
	[0.502]	[0.503]	(0.069)
Missing "advanced"	0.115	0.091	-0.023
	[0.320]	[0.289]	(0.044)
English proficiency	0.482	0.475	0.004
	[0.503]	[0.503]	(0.081)
Missing "English"	0.115	0.091	-0.023
	[0.320]	[0.289]	(0.044)
Class size	50.677	51.011	1.047
	[17.694]	[30.746]	(3.236)
STEM class	0.708	0.682	0.045
	[0.457]	[0.468]	(0.031)
Rural	0.553	0.574	0.009
	[0.243]	[0.234]	(0.020)
Female	0.550	0.545	-0.015
	[0.159]	[0.151]	(0.017)
Minority	0.321	0.321	-0.004
	[0.248]	[0.245]	(0.019)
Age	0.847	0.850	0.002
	[0.105]	[0.094]	(0.011)
Repeater	0.086	0.114	0.034
	[0.243]	[0.256]	(0.036)
CEE score	-0.002	-0.052	-0.071
	[0.780]	[0.716]	(0.079)
F test			0.958
(P value)			0.478
Schools	23	25	
Teachers	96	88	

Notes: This table reports the balance checks results between the control and treated teachers using teacher-level covariates in 2017. Strata fixed effects are included. Joint F test results from Table 2.15 are reported. Standard errors in parentheses are clustered at teacher level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.14. Balance checks for teacher interventions: Student-level covariates

	Control (1)	Treatment (2)	T-C diff (3)	Not in RCT sample (4)
Rural	0.552	0.581	0.005	0.559
	[0.497]	[0.493]	(0.016)	[0.497]
Female	0.543	0.546	-0.015	0.544
	[0.498]	[0.498]	(0.011)	[0.498]
Minority	0.333	0.354	-0.008	0.315
	[0.471]	[0.478]	(0.012)	[0.464]
Age	0.851	0.858	0.002	0.868
	[0.356]	[0.349]	(0.010)	[0.339]
STEM	0.742	0.714	0.044*	0.667
	[0.438]	[0.452]	(0.026)	[0.471]
Repeater	0.118	0.164	0.025	0.229
	[0.323]	[0.370]	(0.035)	[0.420]
CEE score	0.070	0.018	-0.066	0.120
	[0.969]	[0.939]	(0.064)	[0.964]
F test			1.150	
(P value)			0.362	
Students	4,865	4,489		44,701
Schools	23	25		61
Teachers	96	88		962

Notes: This table reports the balance checks results between the control and the treated teachers using student-level data in 2017. Strata fixed effects are included. Joint F test results Table 2.15 are reported. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.15. Balance checks for teacher interventions: Prediction of treatment status

	Treatment group (=1)	
	(1)	(2)
A. Student-level covariates		
Rural	0.005	
	(0.016)	
Female	0.004	
	(0.008)	
Minority	-0.004	
	(0.018)	
Age	-0.005	
	(0.012)	
STEM	0.287*	
	(0.157)	
Repeater	0.084	
	(0.114)	
CEE score	-0.026	
	(0.026)	
B. Teacher-level covariates		
Experience (years)		-0.011
		(0.013)
Missing experience		-0.096
		(0.313)
Monthly salary		-0.000
		(0.000)
Missing "salary"		-0.371
		(0.455)
College choice is important		0.143
		(0.124)
Advanced certificate		0.096
		(0.161)
English proficiency		-0.062
		(0.142)
Class size		-0.000
		(0.002)
STEM class		0.350*
		(0.189)
Rural		0.061
		(0.601)
Female		0.164
		(0.521)
Minority		-0.205
		(0.499)
Age		-0.646
		(0.925)
Repeater		0.478**
		(0.233)
CEE score		-0.090
		(0.135)
Constant	0.261**	1.217
	(0.126)	(0.865)
F test	1.150	0.958
(P value)	0.362	0.478
Observations	9,354	184
R-squared	0.116	0.179

Notes: This table reports the balance checks results from separate OLS regressions that predict the treatment status using both teacher-level and student-level data in 2017. Strata fixed effects are included. Joint F test results are reported. Standard errors in parentheses are clustered at high school (teacher, in column 2) level.* significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.16. ITT and TOT effects of individualized advising programs: Itemized outcomes in the index measure

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

Notes: This table reports the OLS regression results of the ITT effects (Model (2.2)) and TOT effects (Model (2.4)) of the individualized advising interventions in 2017 on itemized college match outcomes (except for college quality in column 6 of Table 2.4) that build the single index. The sample includes the universe of public high school graduates in Ningxia in 2017 (N=54,055). **Machine learning** intervention used the predicted probabilities of admissions to each college-major-order for each student and other automatic data analyses to assist the conventional online personalized advising. **Business as usual** intervention provided low-touch, brief college application guidelines and tips to students, which is used as a placebo test that mimics a low-price for-profit consulting service available to students if they are willing to pay for it. All regressions control for student-level covariates and strata fixed effects. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.17. ITT and TOT effects of individualized advising programs: Itemized college application behaviors

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

Notes: This table reports the OLS regression results of the ITT effects (Model (2.2)) and TOT effects (Model (2.4)) of the individualized advising interventions in 2017 on itemized college application behaviors. Results for the rest strategies and preferences are statistically insignificant. The sample includes the universe of public high school graduates in Ningxia in 2017 (N=54,055). **Machine learning** intervention used the predicted probabilities of admissions to each college-major-order for each student and other automatic data analyses to assist the conventional online personalized advising. **Business as usual** intervention provided low-touch, brief college application guidelines and tips to students, which is used as a placebo test that mimics a low-price for-profit consulting service available to students if they are willing to pay for it. See the text for more descriptions of the outcome variables. All regressions control for student-level covariates and strata fixed effects. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.18. Heterogeneous ITT effects of individualized advising programs

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

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

Table 2.19. Time use in college applications

	Hours after applications open					Later than 2 days	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rural	-5.571**	-3.602*	-2.288***	-1.216**	-1.043**	-0.051**	-0.016**
	(2.247)	(2.132)	(0.647)	(0.481)	(0.504)	(0.022)	(0.007)
Female			-2.804***	-2.660***	-2.558***		-0.040***
			(0.458)	(0.462)	(0.441)		(0.006)
Minority			-0.083	-0.663	-0.722		-0.006
			(0.720)	(0.489)	(0.493)		(0.009)
Age			1.188***	1.056***	1.115***		0.010
			(0.398)	(0.363)	(0.369)		(0.006)
STEM			6.198***	6.488***	6.579***		0.063***
			(0.688)	(0.669)	(0.747)		(0.008)
Repeater			3.289***	3.987***	4.248***		0.036***
			(0.780)	(0.923)	(0.965)		(0.009)
CEE score			18.335***	18.065***	17.994***		0.186***
			(0.808)	(0.847)	(1.014)		(0.008)
Constant	60.649***	66.713***	53.695***	52.960***	52.653***	0.588***	0.525***
	(2.166)	(1.975)	(0.970)	(0.745)	(0.715)	(0.022)	(0.012)
Neighborhood FE	No	No	No	Yes	No	No	No
School FE	No	No	No	No	Yes	No	Yes
Observations	39,385	34,989	39,385	39,385	39,385	39,385	39,385
R-squared	0.006	0.003	0.229	0.234	0.239	0.003	0.130

Notes: This table reports the OLS regression (Model (2.1)) results of the rural-urban gap in college application time use, using data from those who submitted college applications in the control group in 2017. **Hours after applications open** is the total hours from the open dates (June 23 for selective colleges and August 1 for non-selective colleges). Students who did not submit their applications were coded as zero hours (results are similar if excluding these students). **Later than 2 days** is a dichotomous variable indicating that a student submitted her application at least two days later than the open date. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.20. Correlations between application time use and outcomes in college choices and admissions

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

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

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

Table 2.21. ITT effects of teacher incentives: Itemized outcomes in the match index

	Detailed outcomes in match index			
	College median (s.d.) (1)	College mean (s.d.) (2)	College min (s.d.) (3)	Ranking (pctl) (4)
Control mean	-0.197	-0.169	-0.920	56.524
Control s.d.	[1.196]	[1.148]	[1.334]	[28.311]
	A. Without teacher covariates			
Teacher incentives	-0.003	-0.001	-0.015	-0.075
	(0.020)	(0.018)	(0.021)	(0.624)
	¡0.023¿	¡0.020¿	¡0.023¿	¡0.563¿
	B. With teacher covariates			
Teacher incentives	0.000	0.003	-0.016	0.071
	(0.019)	(0.017)	(0.019)	(0.607)
	¡0.023¿	¡0.019¿	¡0.022¿	¡0.538¿

Notes: This table reports the OLS regression (Model (2.3)) results of the ITT effects of the pay-for-performance incentives for teachers in 2017 on on itemized college match outcomes (except for college quality in column 6 of Table 2.7) that build the single index. All regressions control for student-level covariates and strata fixed effects. Teacher covariates include both aggregated data of student covariates and teacher characteristics in the pre-treatment survey data, as described in Table 2.13. Standard errors in parentheses (angle brackets) are clustered at high school (teacher) level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.22. ITT effects of teacher incentives: College application behaviors

	Apply to colleges in ML list	College application Index	Strategy General nudge	Strategy Targeting
	(1)	(2)	(3)	(4)
Control mean	0.305	-0.030	-0.029	-0.028
Control sd	[0.460]	[1.001]	[1.003]	[0.994]
Teacher incentives	-0.004	0.008	0.009	-0.009
	(0.010)	(0.028)	(0.028)	(0.024)
	¡0.010¿	¡0.029¿	¡0.030¿	¡0.026¿

	Strategy Special programs	Preference Tuition & quota	Preference Location	Preference Major
	(5)	(6)	(7)	(8)
Control mean	-0.030	0.030	-0.058	0.003
Control sd	[0.978]	[1.012]	[0.977]	[0.999]
Teacher incentives	0.003	0.019	-0.001	0.030
	(0.030)	(0.027)	(0.023)	(0.025)
	¡0.028¿	¡0.026¿	¡0.025¿	¡0.021¿

Notes: This table reports the OLS regression (Model (2.5)) results of the ITT effects of the pay-for-performance incentives for teachers in 2017 on a family of college application behaviors. All regressions control for student-level covariates and strata fixed effects. Teacher covariates include both aggregated data of student covariates and teacher characteristics in the pre-treatment survey data, as described in Table 2.13. Standard errors in parentheses (angle brackets) are clustered at high school (teacher) level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.23. Heterogeneity in the ITT effects of teacher incentives: By student characteristics

	Urban	Rural	Male	Female	Non-minority	Minority	Low achieving	High achieving
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
N (students)	4,059	5,295	4,263	5,091	6,147	3,207	7,170	2,184
Admission	0.015	-0.008	0.004	0.001	0.002	0.004	0.001	0.004
	(0.010)	(0.016)	(0.013)	(0.012)	(0.012)	(0.018)	(0.014)	(0.003)
College match index	-0.009	-0.002	0.001	-0.010	-0.012	0.009	-0.002	-0.011
	(0.020)	(0.020)	(0.022)	(0.018)	(0.020)	(0.026)	(0.017)	(0.011)
College choice index	0.045	-0.020	0.034	-0.013	0.008	0.009	0.011	-0.004
	(0.030)	(0.034)	(0.030)	(0.030)	(0.032)	(0.051)	(0.034)	(0.019)
Targeting strategy	0.016	-0.029	0.001	-0.017	-0.026	0.024	-0.010	0.002
	(0.032)	(0.027)	(0.025)	(0.027)	(0.027)	(0.038)	(0.026)	(0.025)

Notes: This table reports the OLS regression (Model (2.5)) results of the heterogeneous ITT effects of the pay-for-performance incentives for teachers in 2017 on college match index. All regressions control for student-level covariates and strata fixed effects. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

2.13 Additional Descriptions

2.13.1 Measuring College Application Behaviors Using Actual Choice Data

Based on features of the tier-specific applications in the Chinese centralized college admission system, I focus on three sets of strategies. These strategies are expected to capture some of the main application behaviors for a knowledgeable and skillful student. I have also covered these strategies in our interventions from the application guide "textbook", to school workshop, and to personalized advising. This subsection updates that in Chapter I.

The first set 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 I 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. I 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 I recommend students to think about their applications carefully and the number of colleges to apply to is related to the targeting strategies in the second set variables.
- **[Strategy 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 that the college-then-major admissions give each student only one college temporary admission chance. If a student is eventually rejected by a college due to the unmatched of major applications, she will not be considered by other colleges in the same institutional tier and has to move down to lower tiers. In practice, many students only have strong major preferences, but do not understand the need for this strategy to reduce their rejection risks. I construct this variable by calculating the percent of major applications over total available major numbers given the colleges that a student applied to. Sample mean is 69.9%, with a minimum of 16.7% and a maximum of 100%.

- **[Strategy 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 admissions scores than her CEE score. If that student accepts flexible major assignment within that college, then the college will assign her to a major that still has a spot (but that major may not be her interested one). The flexible assignment is actually to increase admission probability by sacrificing major preferences. I construct this variable by calculating the percent of college applications accepting flexible major assignment over the number of applied colleges. Sample mean is 69.2% with a minimum of 0 and a maximum of 100%. The strategy, which I 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 the accurate predictions and decisions. This set of strategies are the key elements of our behavioral interventions as well as the data analysis in a students' college choice and application. Many students do not understand the underlying mechanisms of college admissions that only rank (but not raw score) matters. They naively compare their CEE score in this year with college admissions raw scores, which results in large errors of identifying college types. Students may use different strategies in different tiers, but I 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, I focus on college-level application behaviors, but those choices of majors within each college is also worth exploring in the future research.

- **[Strategy 2.1] Estimated gap (within 0.15 s.d.).** The behavioral rationale is that students should equate their CEE score to admissions 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 CEE score of 500 should then look at colleges with admissions scores around 550 in 2015. If she applied to colleges with admissions scores around 500 in 2015, she would be very much likely to undermatch. The raw scores vary dramatically over the years. Suppose that the raw CEE scores are 600 and 550 for a student ranked 10,000 in 2016 and 2015, if a student with CEE score of 600 in 2016 applied to colleges with admissions scores around 600 in 2015, she would not be likely to be admitted by an undermatched college, but being rejected by all of her applied colleges. I 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.[69] 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

[69]I 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 I use other choices or a summary statistic of these choices.

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.[70]

- **[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. I 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.[71]
- **[Strategy 2.3] Apply to colleges without admissions data in the prior year.** The number of colleges that admit students in one province may change over time. Each year there are "new" colleges for students to apply to. The behavioral rationale is that students need to infer/predict the admissions data in previous years for these "new" colleges using other information, and they may take risks of applying to these colleges. However, if most students are risk-averse and do not apply to those colleges, it is a good opportunity for skillful students to gain an overmatched admission. I construct this variable by identifying students who applied to colleges in the match tier without admissions data in the prior year. Sample mean is 2%.
- **[Strategy 2.4] Descending rank-order list 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,

[70]Figure 1.8 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.

[71]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.

and to minimize the risks of being rejected by all (Hoxby and Avery, 2013). In order to correctly identify types of reach, peer and safety colleges, students need to understand the classification of these types (a rule of thumb is a 0.05-0.15 s.d. threshold) based on score-equating. Then, for the four college choices within each tier, given the institutional feature of Differed Acceptance (Parallel) mechanism, students should list their four choices in the descending order (choice A > choice B > choice C > choice D), otherwise any choices in higher orders with higher *ex post* admissions scores are meaningless. I construct this variable by a dichotomous indicator of students who did so in their match tier. Sample mean is 31%.

- **[Strategy 2.5] Targeting.** The behavioral rationale is that, although students are nudged to apply to a mix of reach, peer and safety colleges, they should not aim too high or too low. In other words, they need to have a tight range of colleges (centering around their CEE scores). I 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 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 information and knowledge to differentiate/understand different AA programs. National AA programs are of high quality (in selective colleges), but provincial AA programs are lower-quality. I construct this variable by identifying that if a student applied to any AA programs. Sample mean is 22%, with a minimum of 0 and a maximum of 1.

- **[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. I construct this variable by identifying that if a student applied to any early admissions programs. Sample mean is 15%, with a minimum of 0 and a maximum of 1.
- **[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. I construct this variable by counting the percent of applied majors in teacher's education. Sample mean is 5.2%, with a minimum of 1 and a maximum of 40.

Student preferences and tastes are individual-specific strictly unobservable. Particularly in constrained college applications, revealed preferences may not be precisely true. I 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 college with larger admissions 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 admissions probability (Kamada and Kojima, 2015), but students may be unaware of the quota information, which is provided to them by the Department of Education. I 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 (Hillman, 2016; Hoxby, 2000; Long, 2004; Miller, 2017; Ovink *et al.*, 2018). It is also true that high-quality colleges concentrate in the economically developed regions in China. Ningxia province, as a low-income region, lacks high-quality colleges. I construct this variable by calculating the percent of applied colleges locating in out-of-province regions (excluding economically advanced regions and Ningxia's neighborhood provinces, the latter is treated as "in-province"). Sample mean is 38.8%, with a minimum of 0 and a maximum of 1.
- **[Preference 2.2] Out-of-province (advanced regions) colleges.** underlineI construct this variable by calculating the percent of applied colleges locating in the most economically advanced regions of China, including Beijing, Shanghai, Guangdong. Sample mean is 6.6%, with a minimum of 0 and a maximum of 1.

The last set of preferences are major choices. I 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.** I construct these variables by calculating the percent of each major group over the total number of applied majors. The mean values of those majors in Economics-related, Agricultural-related, Computer science-related, International-related, and Medical-related are 24.1%, 1.3%, 3.2%, 1.6%, 11.4%. I did not provide direct interventions on major choice but provided information about all the majors (e.g., coursework, college life, labor market outcomes). I 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).

2.13.2 Intervention Descriptions (One-on-One Advising)

Advising Work-Flow

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

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

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

programs, and college-major trade-offs. The decision would be based on the predictions in Step 6.

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

How does machine learning work?

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

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

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

2.13.3 Intervention Descriptions (Teacher Pay-for-Performance)

关于做好 2017 年高考志愿填报指导工作的通知

各县（市、区）教育考试中心：

目前，在高考填报志愿的过程中，大部分考生可能会面临接收到的志愿填报信息不完整、缺乏较为有效志愿填报指导的问题，“高分低报”的现象时有发生。为做好对考生志愿填报服务工作，我院联合北京大学、宁夏大学合作开展我区 2017 年普通高考志愿填报指导的试点工作。试点工作在对我区各普通高中学校进行科学严谨的调查研究的基础上，对 2017 年抽样范围内的高考考生提供志愿填报的指导服务，协助考生实现大学选择和专业选择的最优决策，进而提高高考成绩与志愿的匹配，充分反映我区高中学校教育教学的成效。

为确保此项试点工作顺利进行，现就有关事项通知如下：

一、问卷调查。

请各普通高中学校校长、高三年级组长（级长）、和全体班主任，务必在 2017 年 6 月 5 日 24 时前，登录北京大学网络问卷调查系统（6 月 1 日 0 时上线），准确完成相应问卷全部内容的填写。

网站登录办法：

（1）电脑浏览器输入：http://surveyor.pku.edu.cn/limesurvey/index.php?sid=83188&lang=zh-Hans;

（2）电脑浏览器 http://surveyor.pku.edu.cn/limesurvey/，点击进入“高考志愿填报信息服务学校调查问卷”；

（3）手机微信或 QQ“扫一扫”附件的二维码。

二、高考志愿填报服务指导。2017 年 6 月 8 日高考结束之后，我院联合北京大学、宁夏大学将通过问卷调查中记录的联系方式，直接联系各高中学校，通过高中学校向抽样考生提供高考志愿填报服务指导。

三、工作要求。请各县（市、区）教育考试中心务必及时通知到辖区内各普通高中学校，按照上述要求，及时完成线上问卷调查，配合我院确保后续各项服务指导工作有序开展。

宁夏教育考试院

2017 年 5 月 24 日

附件：

北京大学网络问卷调查系统“学校调查问卷”二维码（手机微信或 QQ“扫一扫”即可登录）

(a) Survey annoucement by Ningxia DOE

北京大学
PEKING UNIVERSITY

北京大学中国教育与人力资源研究中心
“如何填好高考志愿”在线讲座

邀请信

亲爱的老师，您好！

感谢参加北京大学中国教育与人力资源研究中心面向 2017 届宁夏自治区高三毕业班师生开展的“高考志愿填报公益帮扶项目”！感谢您及时、认真地完成我们的前期调查问卷，您提供的信息有助于我们研究全区高考志愿填报的现状，以期对后续的志愿填报信息服务和政策决策提供科学依据。我们接下来进行的一项重要工作是**帮助高中老师为学生提供高考志愿填报指导**。

学生因为志愿填报不得当而产生的“高分低报”、“高分低录”问题，使得这些学生错失了接受更好学校教育的机会。究其原因，这些学生在填报志愿时普遍缺少有效的信息和科学的决策方法，包括（1）不清楚高考志愿的相关政策（如平行志愿、提前批、专项计划等）；（2）缺乏有关不同档次、类型高校及专业的信息；（3）不清楚自己的院校、专业兴趣；（4）缺乏科学的志愿填报方法（如过于求稳或冒进）；等等。

高中是学生填报高考志愿时非常重要的信息来源，对贫困学生而言可能是最为重要的或者唯一的信息来源。高中老师一般具有科学填报志愿的能力，较为丰富的辅导经验，以及掌握了及时准确的相关信息。同时，同一班级内部，相似情况的学生集中，有利于实现老师同时辅导多个学生的规模效应。

鉴于此，**我们强烈建议您为您所在班级的学生提供必要的一对一咨询和答疑活动。**我们的研究团队在 2016 年的试点项目发现，为学生提供志愿填报相关信息和一对一辅导，将非常显著地提高学生的录取结果。我们认为您在接下来一周的工作，可以使得您的学生，在高考分数既定的情况下，通过良好的志愿填报，更上一层楼！

为了协助您完成这项意义深远的工作，以及回答老师们普遍面临的问题，**我们诚挚地邀请您参加即将到来的“如何填好高考志愿”在线讲座**：

讲座时间：2017 年 6 月 22 日（周四），下午 3-5 点或晚上 7-9 点（任选一场）

主讲人：丁延庆教授，博士生导师，北京大学教育经济与管理系主任
叶晓阳，美国密歇根大学博士候选人，北京大学本科、硕士

讲座内容：（1）高考志愿三问（想上什么学校？能上什么学校？该报什么学校？）
（2）理解高考录取的关键规则（平行志愿、专业录取与调剂）
（3）志愿策略的选择（批次、冲保、专项）

参与方式：
请添加我们的服务微信“北京大学专家组-高考志愿辅导”（微信号：gkto201703）为好友，工作人员将会与您沟通更多详情。为鼓励您的参与，我们准备了微薄的现金红包；我们同时也需要排除没有被邀请的听众。

期待您的参与！感谢您对宁夏自治区高中教育的辛勤付出与贡献！

北京大学教育学院
北京大学中国教育与人力资源研究中心
2017 年 06 月 20 日

地址：中国 北京市海淀区颐和园路5号
邮编：100871
www.pku.edu.cn

(b) Information letter

Figure 2.12. Teacher pay-for-performance information letter in 2017

Notes: This figure shows two information letters I sent to teachers before the official pay-for-performance contract. The first letter (panel A) was sent directly by the Ningxia Department of Education to all 12th grade class head teachers, introducing the project in general and inviting teachers to complete the online survey. The second letter (panel B) was sent by our research team, using the valid contact information from the survey. It introduced more about undermatch and the potential role of high school teachers in helping students. Both treatment and control teachers were invited, but the take-up was low.

北京大学
PEKING UNIVERSITY

北京大学中国教育与人力资源研究中心
“班主任辅导学生志愿填报”绩效奖励计划

亲爱的老师，您好！

感谢参加北京大学中国教育与人力资源研究中心开展的“高考志愿填报公益帮扶项目”！感谢您及时、认真地完成我们的前期调查问卷，您提供的信息有助于我们研究全区高考志愿填报的现状，以期对后续的志愿填报信息服务和政策决策提供科学依据。作为其中一项重要的工作，我们希望帮助高中的班主任老师为学生提供高考志愿填报指导。

我们在 6 月 22 日和 23 日举行了两次“如何填好高考志愿”在线讲座，有的老师可能因为时间关系没有来得及参加。我们现在将讲座的录像视频上传到了网上，如果您感兴趣的话，欢迎随时观看，有任何的问题，我们很高兴予以沟通和解答。

北京大学丁延庆教授高考志愿讲座”
视频地址 1（优酷网）：
http://v.youku.com/v_show/id_XMjg0NjM1NDIyMA==.html?spm=a2h3j.8428770.3416059.1
视频地址 2（Bilibili 网）： http://www.bilibili.com/video/av11582407/

现在正是学生紧张填报高考志愿的时候，我们相信老师您也在忙碌于帮助学生做出人生的一个重要选择！为了对老师们的辛勤工作提供激励，北京大学中国教育与人力资源研究中心从国家自然科学基金筹集经费，为随机抽取的一部分老师提供绩效奖励。**通过问卷调查，我们非常高兴地为您提供“班主任辅导学生志愿填报”绩效奖励计划（请见下一页）。**

如果您接受我们的绩效奖励计划，请您通过以下任何方式之一与我们确认：
1) 回复这封邮件；
2) 添加项目联系微信号（发送您的姓名和所在学校）：gkto201703；
3) 发送您的姓名和所在学校到：133-1102-1471

如果您有任何问题，欢迎您与项目负责人联系：

丁延庆，副教授、博士生导师
北京大学中国教育与人力资源研究中心常务副主任
电话：138-1029-1440，邮箱：yqding@gse.pku.edu.cn

感谢您对我们项目的支持，也对您为高中毕业生志愿填报做出的贡献表示敬意！

北京大学教育学院
北京大学中国教育与人力资源研究中心
2017 年 06 月 24 日

地址：中国 北京市海淀区颐和园路5号
邮编：100871
www.pku.edu.cn

(a) Invitation letter

北京大学
PEKING UNIVERSITY

北京大学中国教育与人力资源研究中心
“班主任辅导学生志愿填报”绩效奖励计划

1. 在 2017 年高考录取工作结束以后，我们将从考试院获取全省/区所有考生的高考分数、志愿填报情况，以及他们最终的录取信息。

2. 根据这些信息，我们将计算他“高考实际分数”与他“最终被录取学校-专业的提档分数线”之间的差额，并以此衡量该考生志愿填报的合理程度：高考总分越接近提档分数线则说明志愿填报越有效率，反之则越无效率；高考总分越高的学生如果发生没有被高校录取的情况，则说明志愿填报越无效率。

3. 在考虑到不同考生的高考分数的差异的基础上，我们将使用上述指标对全省/区所有班级进行一次大排名，以反映哪些班主任老师为考生提供了更加有效的志愿填报支持。

4. **如果您所在的班级在这一排名中处于全省/区前 30%的位置，北京大学中国教育与人力资源研究中心将向您颁发“高考志愿辅导优秀教师”奖牌，并向您个人支付 3 千元人民币的现金奖励；如果您所在的高中在这一排名处于前 5%的位置，对您的个人奖励金额将提高至 5 千元人民币。**

5. 根据与考试院的数据协议，我们公布结果及发放奖励的时间预计在 2018 年 2 月。我们将根据您在问卷里留下的电话和邮箱与您联系，如有变动，请及时告知。

北京大学教育学院
北京大学中国教育与人力资源研究中心
2017 年 06 月 24 日

地址：中国 北京市海淀区颐和园路5号
邮编：100871
www.pku.edu.cn

(b) Letter appendix: Details of the performance pay contract

Figure 2.13. Teacher pay-for-performance contract letter in 2017

Notes: This figure shows the formal pay-for-performance contract letter I sent to treated teachers on 24 June, 2017. The first page (panel A) announced the recipient's eligibility and contact information for her to opt in (all the 88 treated teachers chose to opt in), as well as Internet links to the mini-lectures of the training sessions. The second page (panel B) introduced the performance pay contract in detail.

CHAPTER III

Religion and Motivated Cognition: When Ramadan Meets the College Entrance Exam

3.1 Introduction

Classical economic theories generally model that people could correctly update their beliefs as Bayesians. Yet in reality, it is prevalent and costly that many people actually fail to learn from objective information on important issues. For examples, at the macro level, public opinions are polarized on topics such as global warming (Hart and Nisbet, 2012), GMO foods (Priest, 2000), and evolution (Plutzer and Berkman, 2008), despite overwhelmingly one-sided scientific evidence; at the micro level, individuals fail to process red flags of disease risks (Oster *et al.*, 2013), and traders fail to internalize clear signals of housing-market crashes (Cheng *et al.*, 2014). In education, while there is rapidly growing evidence on the informational interventions that improve behaviors and educational choices for low-income and underrepresented minority students, informational interventions are ineffective when those disadvantaged students do not learn from the information and take actions (Lavecchia *et al.*, 2016; Page and Scott-Clayton, 2016).

To explain such learning failure, one of the theories proposed by behavioral economists is that people have motivated beliefs (Bénabou, 2015; Bénabou and Tirole, 2011, 2016): they attach psychological values to certain beliefs, and can distort their own perceptions or understanding toward such preferred beliefs to increase their utility.[1] However, to fully explain the "learning failure" in reality, existing evidence still falls short on two important dimensions. First, while a series of lab experiments have demonstrated the potential of this theory,[2] investigation in the field has met with considerable challenges due to the difficulty in exogenously varying people's psychological motivation. Second and more importantly, while existing works have shown that people can distort subjective opinions in their own minds, there is little evidence on the stronger form of motivated beliefs: whether/how people can distort objective information, even when it is directly presented in front of their eyes. Such distortion of credible signals is referred to as "reality denial" by Bénabou and Tirole (2016), which is a key implication that distinguishes the theory of motivated beliefs from other theories explaining belief distortion.

Motivated by these gaps in our knowledge, in this paper, we combine a large-scale high-stakes natural experiment with a novel field experiment to provide the first piece of empirical evidence on how people can distort high-stakes objective information in a real setting. In particular, we focus on a setting where religious practices conflict with reality, and reveal a specific underlying mechanism for information distortion: the stringency of religious constraints leaves religious followers no other choice but to obey religious rules, motivating them to self-rationalize their own religious behaviors to avoid regretting such decisions. This means they would distort objective signals suggesting high costs of their religious behaviors.

Our natural experiment comes from a unique empirical setting in China. In this context, China's College Entrance Exam (CEE), which is the single criterion for college admission for almost all the high school graduates and is perceived to be of extremely high stakes (Jia and Li, 2016), is held on June 7th and 8th every year. Between 2016 and 2018, the exam

[1]This literature also closely relates to an older psychology literature on motivated reasoning, as summarized by Kunda (1990).

[2]For examples, Eil and Rao (2011) and Mobius *et al.* (2011) show that people exhibit asymmetric updating behavior about self-image, and Di Tella *et al.* (2015) show that beliefs about others' altruism decrease with stakes.

happened to fall in the month of Ramadan, which follows the Islamic Lunar Calendar and shifts 11 days forward every year in the Gregorian Calendar. Using a difference-in-differences estimator with administrative data on all students of high school graduation cohorts between 2011 and 2016, we find that taking the exam during Ramadan has a quantitatively important negative impact on the exam performance of Muslim students, as compared with their non-Muslim counterparts.

Given the large negative impacts of Ramadan fasting on exam performance, and the extremely high stakes associated with the CEE score, it would be natural for the Muslim students who were about to take the CEE during Ramadan in 2018 to consider the possibility of breaking the fast for the exam. However, for those who were practicing Muslims, whether or not they could delay the fast until after the exam depended on obtaining an exemption from a local expert in Islamic jurisprudence (Faqih), explicitly allowing them to break the fast for the CEE. Seeing that there was no clear exemption from fasting made public for the CEE in our setting, we invited two well-respected Muslim religious leaders to explicitly grant such exemptions (and present the Quranic reasoning) to the Muslim students. Receiving such an exemption gives a student the flexibility to choose whether or not to fast during the CEE, which therefore is a relaxation of the religious constraint.

We conducted a randomized controlled experiment in a large Muslim high school in the Ningxia Hui Autonomous Region, a Chinese province where 38% of the population is Muslim. By randomly providing the exemptions to some of the Muslim students who were about to take the CEE (during Ramadan) in 2018, we created experimental variation in the stringency of religious practices: some students believed that they had to fast during the exam, while other students thought that the fast could be delayed until after the exam. We then showed students the same graph displaying what we found in the administrative data: taking the exam during Ramadan (fasting) has salient negative impacts on the exam performance of Muslim students. Using a visual-based survey module, we find that students who thought they had to fast during the exam in 2018 (control group) show patterns of motivated cognition: they distort the objective information in the graphs by underestimating the negative impacts of taking the exam during Ramadan, but as much as 40% of such cognitive bias is eliminated among the students who received an exemption (treatment group).

Our baseline findings suggest that students distort objective information on the negative consequences of their religious behavior. We conduct a series of additional analyses to better understand the mechanisms. First, we find that the baseline results are mainly driven by extensive margin effects: Muslim students either became particularly accurate in information acquisition after receiving the exemption, or do not adjust their beliefs at all. This is driven by the fact that the incentives to manipulate beliefs arise from one's fasting behavior during the CEE, which is itself a binary decision. Second, we show that the bias in information cognition is most salient among students who always strictly followed the Ramadan fasting requirements during high school, and they are also the ones who respond strongly to our provision of exemptions. This suggests that the baseline findings are indeed driven by motivations associated with fasting attitudes. Third, using a placebo test, we find that the provision of an exemption to delay the fast does not affect the acquisition of information that is unrelated to the Ramadan fasting. Fourth, using a list experiment approach, we provide suggestive evidence that alleviating motivated cognition makes students better informed about the costs of Ramadan, and thus more willing to delay the fast until after the CEE.[3]

Our paper speaks to four strands of literature. First and foremost, the paper provides a direct and strong test for the theoretical framework of motivated beliefs/cognition (Bénabou, 2015; Bénabou and Tirole, 2011, 2016). Originating from research on motivated reasoning in psychology,[4] the existence and implications of motivated beliefs/cognition have attracted much scholarly interests in both experimental economics and political science[5]. Our study

[3] A list experiment requires respondents to provide the total number of items on a list to which they answer affirmatively, rather than to answer each item separately, alleviating the bias in eliciting answers to sensitive questions.

[4] See Kunda (1990) for an extensive discussion of the psychology literature.

[5] In economics, Eil and Rao (2011); Mobius *et al.* (2011) study the impact of self-image related matters on belief updating. In political science, Redlawsk (2002) study the effects of motivated reasoning on po-

differs from the existing evidence from lab experiments in two important ways. As a field experiment,[6] our setting is based on a real-life event with extremely high stakes, which allows us to study motivated cognition in an important environment outside of the lab. Moreover, lab experiments have shown that subjects can either distort subjective beliefs (Eil and Rao, 2011; Mobius *et al.*, 2011) or selectively acquire information in the presence of substantial cognitive cost (Ambuehl, 2017). Our paper, instead, tests a stronger form of motivated belief: when clear objective (rather than subjective) information is presented right in front of one's eyes (so that there is minimal cost of information acquisition), individuals still can distort such objective information and fail to learn. Bénabou and Tirole (2016) refer to this type of behavior as "reality denial," which is a particularly salient type of bias due to motivated beliefs, yet there has been no empirical evidence to date examining its existence and underlying mechanisms.

Second, our paper builds on a long-standing behavioral economics literature on cognitive limitations (Simon, 1955; Conlisk, 1996). A classic example of cognitive limitation is visual bias, which has been shown to affect information acquisition and could potentially distort economic decisions (Kahneman, 2003). Most of the existing works attribute visual bias to individuals' inherent limits in visual perception. Our paper differs from that literature by showing that visual bias could also arise when individuals are motivated to avoid undesirable graphical information. Specifically, we design a novel visual-based survey module, which, combined with our experimental intervention, allows us to quantify the magnitude of motivated visual bias in information acquisition; to our knowledge, this is new to the literature. In a related paper, Exley and Kessler (2018) find that people make simple mistakes (e.g., trivial calculation errors) when they are motivated to do so. Our paper and their

litical decision-making; Taber and Lodge (2006) study how people process arguments on important public policy issues with different prior attitudes; Nyhan and Reifler (2010) demonstrates that motivation affects information processing in reading news articles about politicians.

[6] More precisely, according to the classification of field experiments in Harrison and List (2004), our experiment can be defined as a framed field experiment in that the subjects of interest (i.e., students who will face the choice regarding fasting on exam day) is precisely the population with whom we are conducting the experiment. Moreover, the intervention we introduce is of critical relevance to their important decision in real life.

paper complement each other as they document different types of motivated distortion of objective outcomes (visual bias vs. computational error), but our paper focuses more on the "information acquisition" aspect of motivated bias. Our visual-based survey module is also valuable in itself: it could be easily generalized to other settings due to the pervasive role of graphs in information dissemination, decision-making and persuasion.

Third, our paper contributes to the literature on religious participation. Existing literature on this topic mainly follows a "rational choice" approach: the decision to participate in religious activities is based on a cost-benefit analysis (Azzi and Ehrenberg, 1975; Iannaccone, 1992, 1998; Montgomery, 1996; Stark and Finke, 2000; Berman, 2000). The validity of such rational choice frameworks critically relies on the condition that individuals can accurately evaluate the costs and benefits of religious participation. On the benefit side, Augenblick *et al.* (2016) find that religious followers sincerely attach high pecuniary values to their religious beliefs. However, since valuation of religious beliefs could come from different motives such as salvation, consumption, and peer pressure, it remains challenging to conclude whether such high valuation of religious benefits is "biased."[7] Our paper complements Augenblick *et al.* (2016) by investigating the cost side of the decision. Since our DiD analysis accurately quantifies the costs of religious participation (taking the exam during Ramadan), we have the unique opportunity to identify the difference between "perceived cost" and "actual cost" of religious participation, and to analyze the driving force of such a discrepancy. Our results suggest that stringent religious constraints define the action of the followers, motivating them to self-rationalize such action by underestimating its potential costs. These findings imply that religious followers do not realize the full costs of their religious behaviors, which could help rationalize prevalent religious participation and rapid religious expansion in the rational choice framework.

Fourth, this paper also adds to an accumulating body of evidence on the impact of food/liquid deprivation on cognitive function and other economic activities. Documenting

[7]See Iyer (2016) for a detailed discussion on the different dimensions of religious benefits.

the negative impact of fasting on exam performance is important to understanding the non-religious implications of religious behaviors practiced by millions of Muslims (Kuran, 2018). There is a growing literature on the adverse effect of fasting during the prenatal period on children's cognitive skill, health and labor supply (Almond and Mazumder, 2011; Almond *et al.*, 2015; Majid, 2015). It has also been demonstrated by Schofield (2014) that Ramadan fasting has serious impacts on cognitive function, and that people may be naive about the determinants of their health status such as caloric intake. A particularly relevant paper is Oosterbeek and van der Klaauw (2013), which estimates the effect of fasting during the semester on final exam performance in an economics course. Our DiD analysis differs from their paper in two important ways. To start with, the exam we study is of much higher stakes. Moreover, the DiD effect we estimate is purely driven by "fasting on exam day," teasing out any differences in learning before the exam.[8]

The remainder of this paper is organized as follows: in Section 3.2, we introduce some background information and present the DiD results on the negative consequences of Ramadan; in Section 3.3, we discuss the details of our experimental design and implementation, and lay out the testable hypotheses which guide the empirical analysis; in Section 3.5, we present the empirical results and discuss underlying mechanisms; in Section 3.6, we conclude.

3.2 Background

In this paper, for both the analysis of administrative data and the survey experiment, we focus on the Ningxia Hui Autonomous Region (henceforce Ningxia), which is a provincial unit with a population of 6.3 million, and has a per capita GDP of $7103 in 2018.

Among the 6.3 million residents in Ningxia, 38% are Hui, a Muslim minority ethnic group in China, and the rest are mainly Han, the majority ethnic group in China (non-

[8]What Oosterbeek and van der Klaauw (2013) identify is a compound effect of these two factors. We can tease out the "learning factor" because in 2016, which is the "treatment period" in our DiD analysis, Ramadan started only one day before the exam.

Muslim).[9] Islam is the dominant religion in Ningxia; there are currently more than 3300 major mosques, and more than 4000 certified Imams; in comparison, there are fewer than 200 religious sites for all the other religions combined, including churches, Buddhist temples, Taoist temples, etc.

Compared with the other major Muslim minority group in China, the Uyghurs, the Hui people are much more similar to the majority Han Chinese: their appearances can hardly be distinguished from the Han people, and their mother tongue is Mandarin.[10] It is generally believed that the Hui have much better relationships with the Chinese government than the Uyghurs, and as a result, the government shows a more lenient attitude towards their religious activities and practices (such as praying and fasting in schools) in the Ningxia Hui Autonomous Region than in the Xinjiang Uyghur Autonomous Region.

In the remainder of this section, we introduce the background of our context: the College Entrance Exam in China, the Muslim Ramadan fasting, and how the overlap between Ramadan and the exam affected the performance of Muslim students.

3.2.1 The College Entrance Exam in China

The College Entrance Exam (CEE) in China is a closed-book written exam held on June 7th and 8th every year. Students take the exam in their province of residence, within which the exam content is the same for students in the same track (Social Sciences or STEM). All students are tested on Chinese, mathematics, and English, each with a maximum score of 150 regardless of their track. In addition, students in the social sciences track take another exam on history, politics, and geography, while students in the natural sciences track take another exam on physics, chemistry, and biology. This track-specific exam accounts for 300 points. The total score of the CEE is therefore 750 points.

[9] According to the Global Religious Landscape Study, Muslims are the second largest religious group (23%, only next to Christians). There are about 10 million Hui Muslims in China, and about one-fourth reside in Ningxia.

[10] In comparison, the Uyghurs speak a Turkic language written with an Arabic script, and their appearance is distinct from the Han Chinese.

All Chinese colleges admit students based on the students' provincial ranking of the CEE score. For the vast majority of students, conditional on their own stated college preferences, provincial ranking of the CEE score is the sole criterion that determines the admissions outcome.[11] Given the tremendous value placed on education in the east Asian culture (Chen *et al.*, 2017), and the huge return to elite college education (Jia and Li, 2016), it is not surprising that the CEE is considered by nearly all parents and students as a life-changing opportunity, and regarded by many as the most high-stakes event in a lifetime.

College Admission in China follows a centralized system, where each student first learns about his own score, then submits a ranked list of preferred colleges, and then the colleges admit students solely based on the submitted lists and exam scores. Due to the highly competitive nature of this matching market, even a modest improvement in CEE score (say 5 points out of 750, approximately 0.05 standard deviation) would typically allow a student to include better colleges in his ranked list, and could easily lead to more desirable admission outcomes. Even within the same college, popular majors such as economics, finance, and computer science are typically only available to the students with higher CEE scores. Therefore, students at any part of the distribution generally have strong incentives to increase their CEE scores, even by just a small margin.

3.2.2 The Muslim Ramadan Fasting

Ramadan is the 9th month in the Islamic Calendar, and is observed by Muslims around the world as the holy month of fasting (Sawm) to commemorate the first revelation of the Quran to Muhammad according to Islamic beliefs. Fasting during Ramadan is regarded as one of the "five pillars (fundamental religious duties) of Islam." It requires abstinence from food and liquids (including water) from dawn to sunset, and is obligatory for practicing Muslims. Exemptions to break the fast are typically made for children, the ill, the elderly, travelers, and breastfeeding women.

[11] Rare exceptions include winners of international Olympiad contests, students who win sports scholarships, students with exceptional art talents, etc.

Many of the conflicts that arise between modern reality and religious practices are not explicitly discussed in the Quran. Under these conditions, practicing Muslims typically rely on a local expert in Islamic jurisprudence (Faqih) to decide whether their cases could be granted an exemption. For instance, the Egyptian national soccer team qualified for the FIFA World Cup in 2018, for the first time in 28 years, but the game was scheduled to start right after the end of the month of Ramadan. Seeing this potential conflict, the Grand Mufti of Egypt, Shawki Allam, granted the Egyptian national squad his permission to postpone their Ramadan fasting obligations in the days building up to the World Cup. On the contrary, the Tunisian national team faced the same problem, but did not get such an exemption, and as a result, the players kept fasting throughout the month of Ramadan while preparing for the World Cup.

Observing Ramadan has the potential to offer individuals some benefits such as feeling closer to God and learning to exercise greater self-control. However, it has also been well-documented by an extensive medical literature that Ramadan fasting negatively affects health, through weight loss, metabolic changes, irritability, headaches, dehydration, sleep deprivation, lassitude, etc. (Hallak and Nomani, 1988; Ziaee *et al.*, 2006; Leiper and Molla, 2003; Lancet, 2009). Not surprisingly, these symptoms caused by Ramadan fasting have been shown to be associated with a feeling of tiredness, loss of concentration, and unwillingness to work (Afifi, 1997; Karaagaoglu and Yucecan, 2000). In the economics literature, Campante and Yanagizawa-Drott (2015) show that Ramadan fasting reduces economic output, and Schofield (2014) documents that agricultural productivity drops during Ramadan, Oosterbeek and van der Klaauw (2013) show that students have lower test scores due to Ramadan fasting. Relatedly, outside of the Ramadan context, Figlio and Winicki (2005) show that schools subject to accountability pressure strategically raise the calorie content of meals on test days in an apparent attempt to boost short-term student cognitive performance.

Due to the difference between the Islamic (lunar) calendar and the commonly used Gregorian calendar, Ramadan shifts 11 days forward every year and has a 33-year cycle.

The detailed fasting schedule changes every year and is different across regions based on each location's latitude, which is publicized locally by the Imams before the start of the month of Ramadan. In some extreme cases, fasting hours could be almost all day long, leaving little time for caloric and water intake.

3.2.3 Ramadan and Exams

Between 2016 and 2018, the month of Ramadan mainly fell in May and June, which were the popular times for final exams and high school and college entrance exams around the world. As a result, millions of Muslim students worldwide faced a dilemma between practicing the Ramadan fasting and excelling in academic exams. For example, as described in an information paper by the Association of School and College Leaders, 2016 was the first time Ramadan had clashed with major exams and tests in the UK since the 1980s, and this overlap will likely continue until 2019/20.[12]

Around the world, school leaders and teachers are making efforts to accommodate and support students during Ramadan, including schools in the United States,[13] the United Kingdom,[14] Germany,[15] France,[16] Canada,[17] and the United Arab Emirates,[18] etc. The popular strategies to help students who fast include rescheduling testing or event times, shortening school days, providing extra and comfortable space, and other accommodations. At the same time, some institutions, such as the Association of School and College Leaders, collected and distributed statements from established Muslim leaders, suggesting that students could delay the fast until after the exam, which is similar in the spirit to the exemptions offered in our treatment group.[19]

[12] "Ramadan: Exams and Tests, 2018", visited on Aug 5, 2018

[13] A K-12 school example from The Seattle Times; a higher education example from USA Today College.

[14] An Op Ed piece at School Week: "How schools can support students during Ramadan?"

[15] An article at World Crunch: "Hungry Students? Postponed Exams? Ramadan in German Schools"

[16] A news report at RT International: "Row over postponing French Muslim students' exams for religious holiday"

[17] A CBC news report: "How are schools accommodating fasting students during Ramadan?"

[18] News report at Gulf News - Education: "Five-hour school days in Dubai during Ramadan"

[19] This intervention from the UK ASCL suggests that, while exemptions are potentially available, many

The problem is particularly serious when students are scheduled to take high-stakes exams and it is not possible to reschedule the exams. For example, Oosterbeek and van der Klaauw (2013) study the effect of Ramadan on final grades of Muslim students in an introductory microeconomics course in the Netherlands, where teaching and exam dates are not adjusted for the fasting period. Using data over five years and a Difference-in-Differences strategy, they find that each one additional week of Ramadan fasting reduces the final grades of Muslim students by about 0.1 standard deviations.

Muslim students in China faced even more serious situations. Between 2016 and 2018, the College Entrance Exam in China, which is fixed on June 7th and 8th, fell in the month of Ramadan. When deciding how they observe Ramadan, students will need to take into consideration (1) the tremendous importance of the CEE for their future, (2) the negative impact of fasting on CEE performance, and (3) any flexibility to delay the fast until after CEE. While there is little doubt that most CEE-takers think highly of the importance of this exam, neither (2) nor (3) are fully clear in the Chinese context: no empirical evidence has been provided regarding the cost of Ramadan on CEE performance, and very little information regarding "whether fast could be delayed until after the exam" could be found on the Chinese internet or other media.[20]

In the following subsection, we first estimate how Ramadan affects exam performance in the absence of any intervention or accommodations, quantifying consideration number (2); in the experimental design to be discussed in Section 3.3, we collect explicit exemptions from Chinese religious leaders and randomly distribute those to some Muslim students, creating experimental variation in consideration number (3).

Muslim students in the UK are likely unaware of this possibility. This is consistent with our anecdotal observation in China, which motivated our experimental design.

[20] Two pieces of relevant information could be found through online search engines: one article written by an Imam arguing that students should keep fasting during the CEE, another a translated piece based on the statement of the Egyptian Grand Mufti, suggesting students could delay their fast under certain circumstances.

3.2.4 The Costs of Taking the CEE During Ramadan

To identify the causal impact of taking the CEE during Ramadan on students' academic performance, we obtain administrative data on the exam performance of every urban student in Ningxia who took the CEE between 2011 and 2016. This information is maintained by the Ningxia Educational Examination Institute, and was the criterion used in the college admissions process. This administrative dataset includes the exam score of every urban CEE-taker in Ningxia during the six-year period, as well as their basic background information, such as ethnicity, gender, age, school, county of origin, etc.[21]

Exploiting the fact that the CEE fell in the month of Ramadan only in 2016, and the fact that Ramadan is expected to affect the performance of Muslim students only, we identify the causal impact of taking the exam during Ramadan by measuring how the Hui-Han gap in exam score changed in 2016, compared with the gaps between 2011 and 2015. As shown in Figure 3.1, the Hui-Han gap in exam score was overall stable between 2011 and 2015: on average Hui students score 15 points lower than their Han counterparts.[22] However, the Hui-Han gap almost doubled in 2016, suggesting that taking the exam during Ramadan had salient negative impacts on the relative performance of Muslim students.

To formalize the graphical patterns, we estimate a simple Difference-in-Differences model:

$$Score_{isct} = \sum_{t \in \{2012-2016\}} \alpha_t \cdot Hui_i \cdot Year_t + \lambda_{st} + \varepsilon_{isct} \tag{3.1}$$

where $Score_{isct}$ is defined as the CEE score of student i, who chooses track s (STEM v.s. Social Sciences), from county c, and takes the exam in year t. Hui_i is a dummy variable that equals 1 if student i is ethnically Hui, and 0 otherwise. $Year_t$ is the year fixed effect, λ_{st} is the track-by-year fixed effect, and ε_{isct} is the error term. Standard errors are clustered

[21] Since we only have data for urban CEE-takers to conduct the DiD analysis, for consistency, the field experiment is also carried out in an urban Muslim high school.

[22] The enlarged gap in 2014 was driven by the fact that more Hui students chose the social sciences track rather than the STEM track, and the social sciences track exam was relatively difficult in 2014. This fluctuation disappears once we control for a Track-by-Year Fixed Effect in the regression analysis.

at the school level to allow for serial correlation within the same high school.

Since the CEE fell in the month of Ramadan only in 2016, we expect the Hui-Han gap in exam scores to be stable between 2011 and 2015. Therefore, for $t \in \{2012 - 2015\}$, α_t should be statistically indistinguishable from zero. Because the Hui students took the exam during Ramadan in 2016, we expect to see a drop in their relative performance, therefore a negative α_{2016}.

Figure 3.2 shows that the regression results are highly consistent with our expectations: α_t is always a precisely estimated zero between 2012 and 2015, prior to the overlap between the CEE and the month of Ramadan, which suggests that the Hui and Han students have parallel trends in exam performance before 2016. In 2016, when Hui students take the exam during Ramadan, their performance drops substantially relative to their Han peers, by a magnitude of more than 13 points, out of the average score of 383. In contrast with Figure 3.1, the pre-trend is flatter since we have controlled for the idiosyncratic fluctuations at the track-year level. The regression results are quantified in Table 3.1, where we also show that the results are robust to the inclusion of gender and county fixed effects, and also robust to collapsing all the pre-treatment data into one control group.

In this context, a score loss of 13 points is a huge burden for the students, and would very likely lead to admission by a lower-ranked college, or at least a " less desirable" major within the same college.[23] It is also worth pointing out that our DiD model estimates an "Intention to Treat (ITT)" effect, rather than a "Treatment on the Treated (TOT)" effect, given the fact that not all Hui students are practicing Muslims, and some of them might not fast during the exam. Therefore, the "real impact of fasting during the exam" would be even larger than 13 points.[24]

[23] To put the magnitude in context, in Ningxia, winning the highly prestigious National Mathematics Olympiad Competition or an international athletics competition would only be rewarded with 5 bonus points in the CEE.

[24] As shown in Table 3.3, in our representative experimental sample, around 54% of high school students never broke a fast, which suggests that the TOT effects could be as large as 24 points (0.24 standard deviations).

3.3 Experimental Design and Implementation

For Muslim students who were about to take the exam during Ramadan in 2018, the huge negative impact of Ramadan on CEE scores in 2016 (as documented in Section 3.2) would likely be perceived as undesirable information. However, correctly understanding this information is of tremendous importance for them, for at least three reasons. First, knowing the cost of Ramadan fasting for exam performance helps them make better decisions about whether or not to delay the fast until after the CEE. Second, knowing this information in advance could help them decide the optimal effort to put into studying for the CEE. Third, this information could help them predict their own exam performance, which could help them make better college choices.[25]

However, if we simply presented our DiD findings to those Muslim students who were about to take the CEE during Ramadan in 2018, the stringent requirements of Ramadan fasting in the Islamic religion could give them psychological motives to discount this undesirable information, and underestimate the cost of Ramadan on exam performance. Motivated by this intuition, we designed and implemented a field experiment in Ningxia in May 2018, which formally tests how the stringency of religious practice (Ramadan fasting requirement) generates motivated cognition regarding the cost of religious behavior (impact of Ramadan on exam performance).

With the assistance of the China Center for Education and Human Resources Research at Peking University, we partnered with a large urban Muslim high school in Ningxia to conduct a survey experiment. The high school is the second largest in its prefecture city, with 24 classes in its senior cohort (about to take the CEE in June 2018). The majority of students are Hui Muslim, and the average CEE score in the school is comparable to the provincial average. More than 80% of the students board at school on the weekdays, making

[25] While the college choices are only made after the announcement of the CEE scores, students need to learn about potentially relevant colleges in advance, based on their expected CEE score. It has been shown that finding a "suitable" shortlist of schools is central to realizing a satisfactory admissions outcome (Ye, 2018).

a student's religious behaviors such as fasting and praying generally observable to other students.

Our survey experiment took place on May 4th, 2018 (about one month before the CEE in 2018), during a 40-minute afternoon class on Friday, simultaneously for the entire senior cohort. The 533 Hui students who were present to participate constitute our population for this study. The students were informed by their classroom head teachers before the study that this is a survey conducted by Peking University, which is our partner institution and the top university in China. Students were also informed that completing the questionnaire could lead to as much as a 20 Yuan cash reward.[26] Given the huge reputation of Peking University among high school students in China, the questionnaires were answered carefully by the vast majority of our subjects, as shown by the fact that most students correctly answered our multiple choice questions based on a 1000-word reading material.

As summarized in Table 3.2, our survey experiment has a 2-by-2 design. Randomly, half of the students received the exemptions to delay for the CEE through reading an article (*Exemption*), the other half read an article on art and philosophy (*No Exemption*). In the meantime, we cross-randomized the information received by students: half of the students were incentivized to read a graph about the Hui-Han CEE score gap (*Information*), while the other half were incentivized to read a graph about Sino-Janpanese income gap (*No Information*). The 533 Muslim students participating in the study were randomized into one of these four arms.

In our treatment reading material, we collected statements from well-respected Chinese Muslim leaders, which directly gave exemptions to students to delay the fast until after the CEE, and combined them as an article of about 1000 Chinese words. Specifically, we interviewed an established Muslim scholar, the Imam of an historic mosque, who explicitly said that "Muslim students should delay their fast until after the CEE is finished." We also interviewed a famous religious leader, who is the vice president of the provincial Islamic

[26] 20 Yuan is not a trivial amount in this setting; an average student spends about 10 Yuan on meals every day.

Association, and were told that "we should interpret the Quran in the modern context and allow the CEE participants to delay their fast." The two Imams also gave the Quranic reasoning behind their arguments. We also collected similar exemptions given in Egypt and France to further support the case. For the control reading material, we edited an article from a famous Chinese writer Xiaobo Wang, which is about different perspectives in appreciating art, and has roughly the same length as the religious reading. For both treatment and control readings, to ensure that students understood the materials correctly, we asked three multiple choice reading comprehension questions after the main texts, and students got monetary rewards if they answered the questions correctly.[27]

Our main outcome of interest is whether a student could accurately acquire the information regarding the costs of taking the CEE during Ramadan. To measure such cognitive accuracy, we presented half the students (randomly selected) in each group (treatment and control) with Figure 3.1, which documented how the Hui-Han gap in CEE score was stable between 2011 and 2015, but enlarged abruptly in 2016. The scale of Figure 3.1 was intentionally labeled in a coarse way, where we only showed the max (0) and min (-40) values, but omitted all the intermediate scales, so that the students had to read carefully to accurately estimate the enlarged Hui-Han gap in 2016.

We explicitly told the students that "*between 2011 and 2015, the CEE did not overlap with Ramadan, and the Hui-Han CEE gap was relatively stable (-14.7 in 2011, and -16.6 in 2015); however, in 2016, the CEE fell in the month of Ramadan, and the Hui-Han CEE gap enlarged in this year. Please read the Hui-Han gap in 2016 from the graph.*" In order to incentivize careful reading of the gap, we offered cash rewards to students whose estimates were in the top 50% in terms of accuracy. The main hypothesis is that if students think they have to fast during the CEE, they would be motivated to underestimate the cost of fasting, therefore students would tend to have downward biases when reading the gap from the graph. On the contrary, when granted the exemption, students think that they do not

[27]The translated versions of treatment and control reading materials can be found in Section 3.10.

have to fast during the CEE, and would thus be able to absorb the information with less psychological motivation, and therefore get more precise estimates from the same graph.

Both our anecdotal knowledge and the recent literature suggest that Muslim students might not be fully aware of the negative impacts of fasting (Kuran, 2018). To verify whether this is true in our context, for the other half of the students in each group, we did not show them the "Hui-Han CEE gap" graph (Figure 3.1). Instead, we just told them "*between 2011 and 2015, the CEE did not fall in the month of Ramadan, and the average Hui-Han CEE gap was -16.4; however, in 2016, the CEE fell in the month of Ramadan,*" and then we asked the students to guess the 2016 Hui-Han CEE gap, in an incentivized way.[28] By doing so, we could elicit students' priors regarding the Hui-Han CEE gap, in the absence of any interventions.

For the students who did not read the "Hui-Han CEE gap" graph, we conducted a placebo test, where we asked them to read a graph on the Sino-Japanese income gap, as illustrated in Figure 3.3. Since exemptions to delay the fast should not affect motivations to distort beliefs about Sino-Japanese income gap, we expect no difference in reading the gap in this graph.

For students in all four arms, in addition to the randomized contents (religious vs. placebo reading; Hui-Han vs. Sino-Japanese information), we also asked them a common set of questions on basic individual characteristics, including age, gender, parental education, access to computer, access to internet, academic track, whether boarding at school, whether the student prays daily, whether the student never broke a fast during high school, etc. We compared the answers to the administrative information maintained by the school to ensure the authenticity of the data.

At the end of the questionnaire, after the students completed all the common questions, the reading materials, and score gap estimation specific to their assigned arm, we also conducted a common "List Experiment" for every student, where we provided five state-

[28]If the accuracy of one's guess is above the median student, he will receive a cash reward of 3 Yuan.

ments about the CEE, four of which were subjective and unrelated to religion, including "*(1) learning alone is more effective than learning in groups, (2) we should care about what we have actually learned more than the CEE score itself, (3) playing sports is good for exam preparation, (4) the CEE mainly tests on the familiarity with the material rather than actual intelligence;*" and one statement was about Ramadan fasting, "*(5) delaying fast until after the CEE is acceptable.*" We asked each student how many of the five statements they agree with, without having to specify which statements in particular. By comparing the number of statements agreed with in each arm, we could estimate the direct effect of relaxing the religious constraint on fasting behavior, as well as the indirect effect of better learning about the cost of religious behavior on fasting attitudes.

Given the 2-by-2 design, we prepared four different types of questionnaires: *No Exemption*No Information*, *Exemption*No Information*, *No Exemption*Information*, *Exemption*Information*. All questionnaires have an identical cover letter explaining that this is a survey conducted by Peking University, and the data is confidential and will be used for purely academic purposes. We pre-randomized the order of the questionnaires before distributing them in each classroom; as a result, the 533 Muslim students were randomly assigned one of the four types of questionnaires. We hired 24 surveyors, each covering one classroom throughout the survey experiment, to make sure that students answered their questionnaires individually, and did not communicate with each other throughout the process. Given that the cover letters were identical and the students did not communicate during the survey, it is most likely that the vast majority of students did not realize that they were assigned differentiated questionnaires until the end of the survey experiment.

To check the quality of randomization, in Table 3.3, we present results from the ANOVA test across the four different arms, for all the baseline characteristics that we collected. As can be seen, the four arms are overall very well balanced with each other, suggesting that the randomization is well-executed. There exists only one case where the ANOVA test is marginally significant (perceived value of college). As will be shown in the tables forthcoming, none of our main results would be affected in any substantial way if we included all these covariates in the regressions.

In addition to checking for balance,Table 3.3 also provides some useful information about our context: less than 5% of students have a parent who graduated from college, consistent with the fact that Ningxia is one of the poorest provinces in China; 85% of the students board at school, which means that whether or not a student keeps fasting is largely observable to his peers; 59% of the students pray every day; and 54% of the students never broke a fast throughout high school, suggesting that there exists substantial variation in the religiosity of students.

3.4 Testable Hypotheses

To rationalize the experimental design and guide the empirical analysis, we propose a simple conceptual framework based on the theory of motivated beliefs. In this model, a subject jointly chooses two parameters: (1) his belief about the average cost of Ramadan on CEE performance; (2) whether or not to break the fast during the CEE. By doing so, he maximizes his own utility, which consists of three components: (a) anticipatory utility of exam results; (b) benefits from sticking to the religious practice; and (c) the cognitive cost of manipulating his own beliefs.

The details of the model, including its setup, mathematical proofs, and formal propositions, are elaborated in Section 3.9. In this section, we simply lay out in words the main testable hypotheses derived from the model, and briefly explain the associated intuitions.

Hypothesis 1 If students are unaware of the negative impacts of fasting on exam performance, reading about the exemption alone does not change such erroneous beliefs.

A conjecture made by Kuran (2018) is that most people who commit to Ramadan fasting are actually ignorant about the potential cost of such behavior. Since our DiD estimate on the "cost of Ramadan fasting on exam performance" is a unique piece of information that

Muslim students could barely have access to, it is expected that in the absence of such objective information, even if they receive the exemption to delay the fast, they will still not update their priors. If this prediction is confirmed, then we can also rule out the possibility that the religious reading alone might contain some information that changes priors.

Hypothesis 2 When reading the 2016 Hui-Han CEE score gap from Figure 3.1, in the absence of the exemption to delay the fast until after the CEE, Muslim students would underestimate the true gap.

This hypothesis is the core implication of motivated beliefs: students who stick to fasting due to religious constraints are motivated to underestimate the cost of fasting. The intuition in our conceptual framework is clear: students face a trade-off between the bad anticipatory utility on exam results due to fasting and the cognitive cost of suppressing the truth in order to be overly optimistic.

Hypothesis 3 When reading the 2016 Hui-Han CEE score gap from Figure 3.1, students who received the exemption to delay the fast would on average get more precise estimates, as compared to those who did not get the exemption.

This is the main test of our paper. Receiving the exemption relaxes the religious constraint, which should alleviate the motivation to underestimate the cost of Ramadan on exam performance, and lead to more accurate reading of Figure 3.1. Note that this is a particularly strong test, as Muslim students were presented with the exact same objective information right in front of their eyes, and they were asked to read the "objective information itself," rather than "how they update their own priors based on such information." Therefore, any difference in graph-reading caused by the exemption should be interpreted as the magnitude of "objective information manipulation," which was motivated by the stringency of religious constraints.

Hypothesis 4 When given the exemption, students adjust at the extensive margin. Either they do not debias at all (non-compliers), or they read the Hui-Han CEE gap accurately (compliers).

Muslim students are jointly choosing "whether to fast during the CEE" and "the belief to hold on the cost of Ramadan." Since the fasting decision is binary, our model would imply that the choice on belief manipulation would also be polarized. This means that adjustment in accuracy happens at the extensive margin: when receiving the exemption, some students do not comply, and remain highly biased in graph-reading; other students comply, and become highly accurate in graph reading.

Hypothesis 5 Students who kept fasting in the past would exhibit more severe bias when reading the 2016 Hui-Han CEE score gap from Figure 3.1, but they would also respond more strongly to the exemption and reduce their bias more.

For students who never broke a fast throughout high school, the religious constraint is likely to be more stringent, which means they would have stronger motivation to underestimate the cost of Ramadan on CEE performance, in order to gain psychological relief. Therefore, students who always fasted should have stronger baseline bias in reading the Hui-Han CEE graph. Meanwhile, these "always-fasting students" also had more binding religious constraints, which means that the exemption to delay the fast for the CEE would have a stronger debiasing effect on them, rather than on those students who did not always fast anyway.

Hypothesis 6 Students with higher valuation for college education would exhibit more severe bias when reading the 2016 Hui-Han CEE score gap from Figure 3.1, but they would also respond more strongly to the exemption and reduce their bias more.

This follows a similar logic to Hypothesis 5. When students value college education more, the higher stakes give them more motivation to underestimate the cost of Ramadan on CEE performance, to gain psychological relief. But in the presence of an exemption, since the information on the "cost of fasting" is more valuable to them,[29] they would respond to

[29]The information could help them maximize their performance in the CEE, which they value highly.

the exemptions more strongly and debias more.

Hypothesis 7 When reading the Sino-Japanese income gap from Figure 3.3, whether or not a student received the exemption would not affect the accuracy of the estimate.

This is the main placebo test of our paper. Since the Sino-Japanese income gap is unrelated to either Ramadan or the CEE, any motivated beliefs associated with reading Figure 3.3 would not be alleviated by the exemptions.

Hypothesis 8 Both getting the exemption and reading the Hui-Han CEE gap could make Muslim students more likely to delay the fast until after the CEE. Importantly, these two interventions are complementary: providing exemption and information at the same time could most effectively persuade Muslim students to delay the fast until after the CEE.

This hypothesis links our interventions to students' willingness to delay their fasts. Exemptions mechanically make students more willing to delay the fast, as they are told to be allowed to do so. Information on the Hui-Han CEE gap (Figure 3.1) could also make students more willing to delay the fast, as they become more aware of the potential costs. Our model predicts that these two interventions are complementary to each other: receiving the exemption relaxes the religious constraint, which alleviates motivated cognition, helping the students better absorb the undesirable information on the negative impact of taking the CEE during Ramadan. Therefore, when we combine the *Exemption* treatment with the *Information* treatment in one intervention, it should have the strongest effect in terms of persuading students to delay their fasts.

3.5 Results

In this Section, we analyze the experimental data, and test each of the hypotheses discussed in Section 3.4 respectively.

3.5.1 Students' Priors on the Cost of Ramadan

We first elicit students' priors on the impacts of taking the CEE during the month of Ramadan. In our 2-by-2 experimental design, there exists an arm where Muslim students receive neither the religious reading (exemption) nor the Hui-Han CEE gap information, which we refer to as "*No Exemption*No Information*." Within this group, students do not get any experimental intervention, so they should maintain their original priors regarding the cost of taking the CEE during Ramadan. Therefore, incentivizing them to guess the enlarged 2016 Hui-Han CEE score gap would elicit their priors on the cost of the Ramadan fast, which is representative for our population due to random assignment.[30]

To help students get a sense of the performance gap in the absence of Ramadan fasting, we informed students about the benchmark: the average Hui-Han gap between 2011 and 2015 is -16.4. Then we incentivized each student to make a guess on the 2016 Hui-Han CEE gap, as accurately as possible. For the 128 students in "*No Exemption*No Information*," the answer we get is -17.9, as compared to the true value of -29.4. This suggests that Muslim students held highly biased priors, and believed that taking the exam during Ramadan has minimal impact on performance.[31]

Given that students are largely unaware of the impact of fasting, Hypothesis 1 predicts that, in the absence of new information, the exemption to delay fasting has a negligible impact on the priors of Muslim students. To test this hypothesis, we compare the elicited guesses on the enlarged 2016 Hui-Han gap between "*No Exemption*No Information*" and "*Exemption*No Information*." Formally, for all the Muslim students who did not read the Hui-Han CEE gap figure (*No Information*), we estimate:

$$Gap_i = \alpha \cdot Exemption_i + X_i' \cdot \beta + \varepsilon_i \tag{3.2}$$

[30] One would argue that there is information about the cost of Ramadan in the very fact that we were asking about the "enlarged gap." However, as shown below, this "information" does not alter students' priors at all.

[31] A T-test suggests that the average guess in this arm is not different from -16.4 in any statistically meaningful way.

where Gap_i is student i's elicited prior of the Hui-Han CEE gap in 2016. $Exemption_i$ is a dummy variable, which equals 1 if student i received the exemption from the religious leaders through our reading materials, and 0 otherwise. X_i is a vector of individual characteristics, and ε_i is the error term. We also define an alternative outcome variable $Deviation_i$, which directly measures the absolute value of how far each student's guess deviates from the true value (-29.4).

Table 3.4 shows that the treatment effect is a precisely estimated zero, suggesting that providing the exemption alone does not change the students' priors on the 2016 Hui-Han gap, confirming Hypothesis 1.

3.5.2 Existence of Motivated Cognition and Effects of Exemption

Hypothesis 2 asserts that Muslim students distort their own beliefs when learning about the cost of taking the exam during Ramadan, which leads to an underestimation of the true cost. Moreover, Hypothesis 3 claims that the relaxation of religious constraint (*Exemption*) would reduce such cognitive bias.

To test these two main hypotheses of our paper, we compare the accuracy of graph-reading between Muslim students in "*No Exemption*Information*" and "*Exemption*Information*." We re-estimate Equation 3.2 for students who were asked to read the enlarged 2016 Hui-Han CEE gap from Figure 3.1. Again, we define an alternative outcome variable $Deviation_i$, which directly measures how far each student's reading deviates from the true value (-29.4).

As shown in Table 3.5, for those without exemptions, the average estimated gap is -24.4, which understates the true gap by about 5 points.[32] When randomly assigned an exemption, the estimated gap enlarged by 2 to 2.2 points, eliminating roughly 40% of the baseline cognitive bias. From column 1 to column 3, it is obvious that the coefficient of interest remains highly robust as we control for class fixed effects and the full set of individual

[32]Compared to the control mean in Table 3.4 (-17.9), the control mean here is -24.4, suggesting that the Muslim students could partially absorb the objective information from the figure, consistent with our model assumption that belief distortion is not costless.

controls, again confirming that the randomization was well-executed. We also get similar results using "Deviation" as the outcome variable: the baseline bias is about 5.8 points, more than 30% of which could be eliminated by the religious intervention.[33]

These empirical patterns confirm the main hypothesis of this paper: the stringency of religious practices leads to motivated cognition regarding the cost of religious behaviors, and the relaxation of religious practice could help alleviate such cognitive bias.

Hypothesis 4 predicts extensive margin adjustment: students who decided not to fast will have highly accurate cognition, supporting evidence for which will be presented in the following subsection. Note that this is not in conflict with our results here that the religious exemption eliminates about one-third, rather than 100%, of the baseline bias in Hui-Han CEE gap estimation, because not all Muslim students would change their fasting behavior after receiving the exemption, and only the compliers would adjust at the extensive margin.[34]

3.5.3 Mechanisms

Extensive Margin Adjustment

Our model implies that when students face a binary choice for the religious practice (to fast or not during the CEE), their choice regarding the extent of belief manipulation is also binary, therefore, the cognition is either highly accurate or much off (Hypothesis 4). The intuition for such polarized equilibria is as follows: if one chooses to keep fasting during the CEE, then it is optimal for him to self-rationalize this behavior by distorting the undesirable information; on the contrary, if one decides to delay the fast until after the CEE, it would be optimal to accurately read the graph, as information distortion is costly.

To visualize whether this prediction is true, in Figure 3.4, we plot the distribution of

[33]The difference between using the two outcome variables arises mainly from the fact that a small proportion of students overestimate the 2016 Hui-Han gap, which gets canceled out in the first but not the second definition.

[34]There are mainly two reasons why students might not comply with the exemption: (1) some non-religious Muslim students might not intend to fast anyway ("always-taker"); (2) some highly religious Muslim students might still decide to fast even with the exemption ("never-taker").

"Deviation" by groups (*Exemption* v.s. *No Exemption*). As can be seen, the treatment group with exemptions tends to have fewer "highly biased" students in the right tail, while having more "highly accurate" students in the left, as compared to the group without exemptions. While this is obviously consistent with our hypothesis that the exemption shifts compliers at the extensive margin, there is also an alternative explanation: students only adjusted at the intensive margin, so students with deviation in the range "6 - 15" shifted to the bin "3 - 6," and students in the bin "3 - 6" shifted to the bin"0 - 3."[35] To rule out this alternative explanation, we conduct a simple balance test: if the "extensive adjustment" hypothesis is true, then the composition of students in the "3 - 6" bin should remain the same with or without the exemption, which should not be the case if the "intensive adjustment" explanation is true. Our balance test finds that in the "3 - 6" deviation bin, all the baseline covariates are orthogonal to exemption, which provides suggestive evidence supporting Hypothesis 4.[36]

The fact that students with exemptions become highly accurate in graph-reading can be quantified using regression analysis. For students required to read the 2016 enlarged Hui-Han gap in CEE score, we define a dummy indicator for the accuracy of their estimation: *Accuracy* = 1, if the estimation is within two points of the true value, and 0 otherwise. We estimate Equation 3.2 using *Accuracy* as the outcome variable, and the results are shown in Table 3.6. It is obvious that students who randomly received an exemption are 15 percentage points more likely to produce a highly accurate estimate, which is consistent with the patterns in Figure 3.4. The coefficient is highly robust to the inclusion of class fixed effects and the full set of individual characteristics.

[35] It is also worth noting that a 15-point deviation suggests blatant reality denial: despite the huge drop in Hui-Han gap shown in Figure 3.1, the students give up potential monetary rewards and insist that the trend remains flat!

[36] Some covariates, such as "access to computers" and "whether always kept fasting," are good predictors of "Deviation." These variables do not differ substantially across treatment arms in the balance test.

Motivation Driven by Fasting History

Hypothesis 5 indicates that those who strictly followed Ramadan fasting in the previous years tend to be more religious, and therefore they should have stronger incentives to manipulate their beliefs to underestimate the cost of Ramadan, and in the meantime should also be more responsive to the exemptions.

In the survey, we asked each student "whether you strictly practiced Ramadan fasting (never broke a fast) throughout high school." Roughly 54% of the students answered "Yes" to this question, and the ratio is balanced across the four arms due to random assignment. To explore the heterogeneity associated with previous fasting behavior and test Hypothesis 5, we estimate the following modified version of Equation 3.2:

$$Gap_i = \alpha \cdot Fasted_i + \beta \cdot Exemption_i + \gamma \cdot Fasted_i \cdot Exemption_i + X_i' \cdot \delta + \xi_i \qquad (3.3)$$

where $Fasted_i$ equals 1 if student i strictly practiced Ramadan fasting during high school, and 0 otherwise. Under this specification, α measures the extra baseline bias of the students who strictly practiced fasting, β identifies the treatment effect of the exemption on students who did not strictly practice fasting, and γ identifies the additional treatment effect of the exemption on students who strictly practiced fasting. According to Hypothesis 5, we expect that $\alpha > 0$, $\gamma < 0$, and $\beta + \gamma < 0$.

As shown in Table 3.7, the negative and significant α indicates that students who strictly practiced fasting had larger downward biases than their non-fasting counterparts. The statistical insignificance and small magnitude of β implies that receiving the religious intervention does not have any meaningful impacts on students who did not strictly practice fasting during high school. The large, negative and significant γ implies substantial heterogeneity in treatment effects across those who strictly fasted and those who did not. Estimates of "$\beta + \gamma$" remain highly significant and robust across different specifications, suggesting that our treatment effect is concentrated among those who strictly practiced Ramadan fasting.

Therefore, the empirical results are highly consistent with our theoretical predictions, further suggesting that the stringency of religious practices, rather than confounding factors, are driving the motivated cognition observed in this context.

Motivation Driven by Valuation for College

Hypothesis 6 states that students who value college education highly would have stronger motivated beliefs, but when provided with exemptions, they would also respond more strongly and debias more. The intuition is that these students care more about the CEE scores, so in the absence of exemptions, they can avoid more utility loss by choosing to believe that fasting is not bad for CEE performance; however, when given exemptions, their higher valuation for the CEE makes them more likely to choose to delay the fast, which leads to more accurate readings of the Hui-Han CEE gap.

To test this hypothesis, we estimate a modified version of Equation 3.3, where we replace the "*Fasted*" dummy with a "*High Stake*" dummy, which measures whether the student self-reports having high valuation for college education. As reported in Table 3.8, in the absence of exemptions, students who report high valuation for college education also have 35 percent (1.7 points) extra baseline bias in reading the Hui-Han CEE gap; but in the presence of exemptions, these high-stake students respond more strongly by fully eliminating the extra baseline bias.

Compared with the results on fasting history reported in Table 3.7, the results in Table 3.8 are in general weaker: while the signs of main coefficients are always consistent with Hypothesis 6, they tend to have smaller magnitudes and are only marginally significant/insignificant at the 10% level. The comparison between Table 3.7 and Table 3.8 suggests that both "motivation due to fasting history" and "motivation due to valuation for college" are driving the baseline patterns of motivated cognition, but the former plays a more salient role.

Placebo Test

As stated in Hypothesis 7, receiving the exemption to delay fasting should not affect the cognitive accuracy regarding topics unrelated to either the CEE or Ramadan fasting, such as the Sino-Japanese income gap (Figure 3.3).

Therefore, to further rule out alternative mechanisms, we conduct a placebo test, where some students read the religious article (*Exemption*) and were required to read the Sino-Japanese income gap (*Exemption*No Information*), and others read the placebo article (about art) and were required to read the same Sino-Japanese income gap (*No Exemption*No Information*). Estimating Equation 3.2 for students that read the Sino-Japanese income gap from Figure 3.3 would therefore estimate the "placebo effect of exemption on cognitive accuracy." Note that being Chinese, these students could very well have their own motivated beliefs regarding the Sino-Japanese gap, but what we focus on here is that such motivated beliefs should not be affected by religious exemptions.

Table 3.9 shows that, students tend to underestimate the Sino-Japanese income gap.[37] But importantly, reading about the religious exemption has no statistically meaningful effect on the accuracy of reading the Sino-Japanese income gap, suggesting that our findings are indeed driven by religion-motivated learning, rather than alternative mechanisms.

Active Information Distortion

Another interpretation of our result is that maybe the religious reading (*Exemption*) did not relax the religious constraint, but simply triggered students' interest/curiosity in the topic of Ramadan, so that they gave more attention to the Hui-Han figure, which led to more accurate readings.

This interpretation is inconsistent with the previous findings, which showed that the baseline bias is driven by fasting history and valuation for college, unless students who never broke a fast and students who valued college more highly were also those with no

[37] The true gap is 30771.29, while the students in the control group on average read -28433.92.

interest/curiosity in Ramadan in the absence of the religious reading.

To further address this concern, we propose another test, where we compare the accuracy of graph-reading across all four arms. Since the Hui-Han figure ranges from 0 to -40, and the Sino-Japanese figure ranges from -25000 to -45000, a 2-point deviation in the former is equivalent in scale to a 1000-dollar deviation in the latter. Therefore, we can extend the definition of "Accuracy" to every student in any of the four arms: it equals one if the student either read the Hui-Han gap and made an error within 2 points, or read the Sino-Japanese gap and made an error within 1000 dollars; and zero otherwise.

Following this definition, we are able to compare the accuracy of graph-reading across the 4 different arms. Since the Hui-Han information is more relevant for those Muslim students about to take the CEE during Ramadan, if the results are indeed driven by attention (curiosity/interest), we should expect *No Exemption*Info* to be more accurate than *No Exemption*No Info* and *Exemption*No Info*. However, as shown in Table 3.10, the students are least accurate when reading the Hui-Han figure without an exemption, even less accurate than when reading the Sino-Japanese gap.

This finding suggests that the motivated bias in baseline was driven by active information distortion rather than lack of attention, which supports our framework and rules out the alternative interpretation.

3.5.4 Fasting Decisions

We have demonstrated that our treatments have substantially changed students' beliefs about the impact of Ramadan fasting on exam performance. As predicted by Hypothesis 8, changes in beliefs should also lead to changes in fasting decisions, which we test empirically in this subsection.

The main difficulty here is that the fasting decision is essentially unobservable, and direct elicitation of fasting attitudes in the survey may raise social image concerns, and therefore lead to biased answers. To circumvent these problems, we use a list experiment, commonly seen in the political science literature, to elicit students' attitudes on whether it is acceptable to stop fasting during the CEE. The list experiment is different from direct elicitation in that it asks students "among the following five statements, how many do you agree with?" In the five statements, one of the statements is about fasting attitudes, which we are interested in, and the other four statements concern ways of CEE preparation which are irrelevant to students' religion, with answers depending on students' preferences. Students only have to tell us how many of the statements they agree with and need not to answer explicitly which statements specifically. To minimize the potential social pressure, we also tell students that their response will not be released. Under these procedures, students should be free to state their attitudes in a roundabout way.[38]

As asserted by Hypothesis 8, when students are generally unaware of the harm of fasting during the exam, both accurate information and the exemption granted by religious leaders may be helpful in terms of changing fasting attitudes: the former increases the cost of fasting, while the latter reduces the return. While the model does not specify which type of intervention is more effective, we do know from the prediction that combining both exemption and information will be most helpful in changing attitudes, as there is an interaction effect between the religious and information interventions: relaxing the stringency of religious norms could lead to better acquisition of information, which further changes religious behavior.

To test these predictions, we first compute the average number of statements that students agree with in each of the four arms. Figure 3.5 suggests that both "information alone (*No Exemption*Information*)" and "exemption alone (*Exemption*No Information*)" are helpful in changing students' fasting attitudes, and religious reading appears to be more effective. Another take-away is that the combination of both religion and information is most effective, which is consistent with the prediction of our theoretical framework. To make sure

[38] The proportion of students who agree/disagree with all 5 statements is extremely low (below 2%), which guarantees the effectiveness of the list experiment in "hiding" the fasting attitude of an individual student.

that our finding is statistically robust, we run the following regression:

$$List_i = \gamma_1 + \gamma_2 Exemp \cdot No\ Info + \gamma_3 No\ Exemp \cdot Info + \gamma_4 Exemp \cdot Info + X_i'\beta + \nu_i \quad (3.4)$$

where $List_i$ is the number of statements that student i agrees to in the list experiment. γ_1 is the constant representing the average of $list$ for students in group "$No\ Exemp * No\ Info$," while γ_2, γ_3, and γ_4 represents the point estimate of different treatment effects relative to this baseline. It follows from Table 3.11 that the statistical evidence we find is consistent with the visual representation, and robust across different specifications. The positive and significant γ_2 represents the mechanical effect of allowing students to delay fasting; the positive yet insignificant γ_3 suggests that the information treatment alone is not as effective as the exemption; whereas the large and significant γ_4 suggests that combining both information and exemption will create the strongest effect in persuading students to stop fasting during the CEE.

The fact that the magnitude of γ_4 is larger than the coefficients for γ_2 and γ_3 combined suggests that, in addition to the direct effects of relaxing religious constraints and receiving information on religious behavior, there is also a more subtle interaction effect, where the relaxation of religious constraints leads to better understanding of the information, which further affects religious behavior. The T-test for this additional interaction effect, however, lacks statistical power due to sample size limitations.

3.6 Conclusion

In this paper, we first document that taking the CEE during Ramadan in 2016 had a large and statistically significant negative impact on the performance of Muslim students in China. We then collected explicit exemptions from well-respected Chinese Muslim religious leaders encouraging students to delay the fast until after the CEE, which we randomly distributed to Muslim students who were about to take the CEE during Ramadan in 2018, creating experimental variation in the stringency of religious practice. After that, we presented all students with the same information regarding the cost of taking the CEE during Ramadan, and find that students who thought they were required to fast during the exam were more likely to distort this undesirable signal, by underestimating the negative impacts of Ramadan on the CEE score of Muslim students; but for those students who were randomly selected to receive exemptions to delay the Ramadan fasting, more than half of such cognitive bias could be eliminated.

Further analysis suggests that the baseline treatment effects are driven by adjustment at the extensive margin (students persuaded by the exemptions interpreted the signal highly accurately), and the baseline bias and treatment effects are both particularly strong for students who strictly practiced Ramadan fasting throughout high school and for students who had higher valuations for college. Reassuringly, our placebo test confirms that religious exemption only affects the cognition of religious information (Hui-Han CEE gap), but not the cognition of non-religious information (Sino-Japanese Gap). Our analysis also suggests that while providing either information or exemption alone could potentially change students' fasting behavior, they are most effective when combined together due to the existence of an interaction effect: the exemption could help students better interpret the information on the cost of Ramadan, leading to more informed fasting decisions (increased willingness to delay the fast for the CEE).

In addition to contributing to the growing literature on motivated beliefs, visual bias, religious participation, and Ramadan fasting, the results in this paper also have essential policy implications. Specifically, our findings imply that the dissemination of accurate information and the relaxation of religious constraints work as strong complements when people appear to be ignorant about the adverse impacts of certain religious practices. Therefore, in order to minimize the cost due to conflicts between religion and reality, if accommodating to the religious schedule is unfeasible, a natural second-best solution is to combine "relaxation of religious constraint" with "powerful reminders of the real-life costs of religious behaviors" as a compound policy instrument. More generally, our findings also suggest that, to increase the policy intervention effectiveness, in addition to providing convincing scientific evidence and objective information, it is also important to identify and tackle the psychological motives that could potentially prevent one from acquiring accurate information.

3.7 Figures

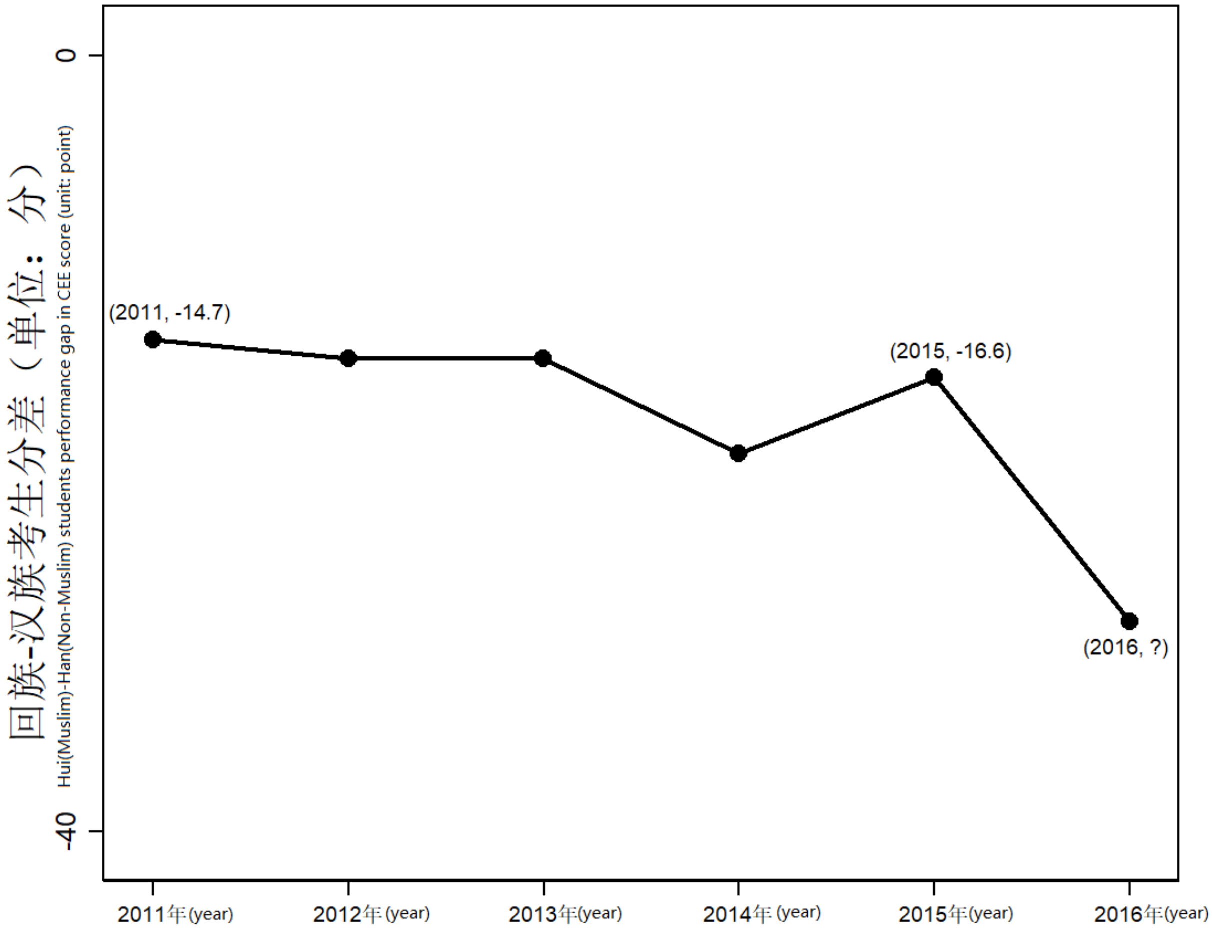


Figure 3.1. Hui-Han score gap in the College Entrance Examination (2011-2016)

Notes: This figure displays the Hui-Han average College Entrance Examination (CEE) score gap for all urban students in Ningxia between 2011 and 2016. This is the same figure that was presented to the students in our experimental sample (with Chinese labels). The first dimension of the coordinates marked beside a data point is year (horizontal axis) and the second dimension of the coordinates (vertical axis) is the magnitude of the gap.

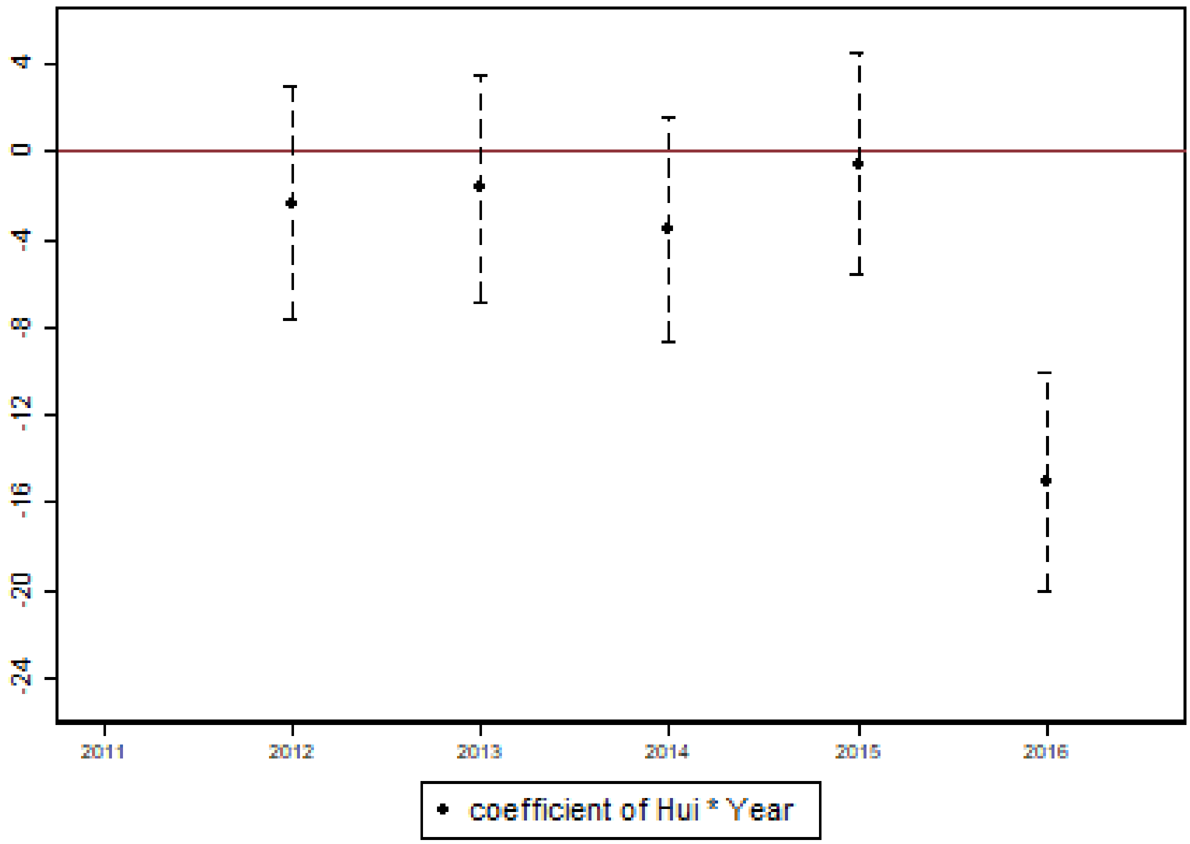


Figure 3.2. Event study estimates of the Hui-Han CEE gap

Notes: This figure presents the event study estimates of the dynamics of the Hui-Han CEE score gap, with 5% confidence intervals plotted around each coefficient estimate (dot). The Hui-Han gap remained highly flat during 2011-2015, but enlarged significantly in 2016, the first year when Ramadan fasting overlapped with the CEE.

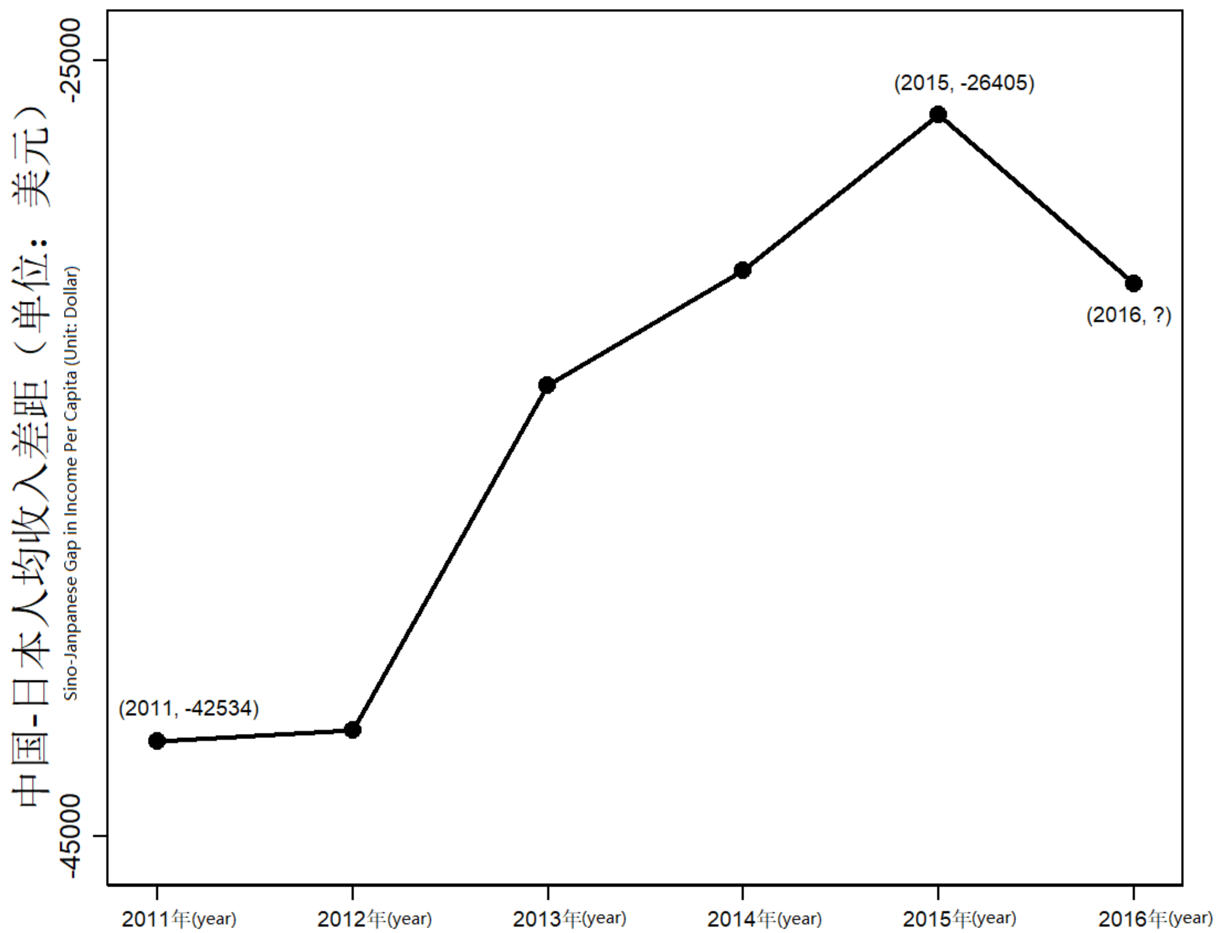


Figure 3.3. Sino-Japanese income gap (2011-2016)

Notes: This figure displays the gap between GDP pc of China and that of Japan during 2011-2016. This is the same figure that was presented to the students in our experimental sample (with Chinese labels).The first dimension of the coordinates marked beside a data point is year (horizontal axis) and the second dimension of the coordinates is the magnitude of gap.

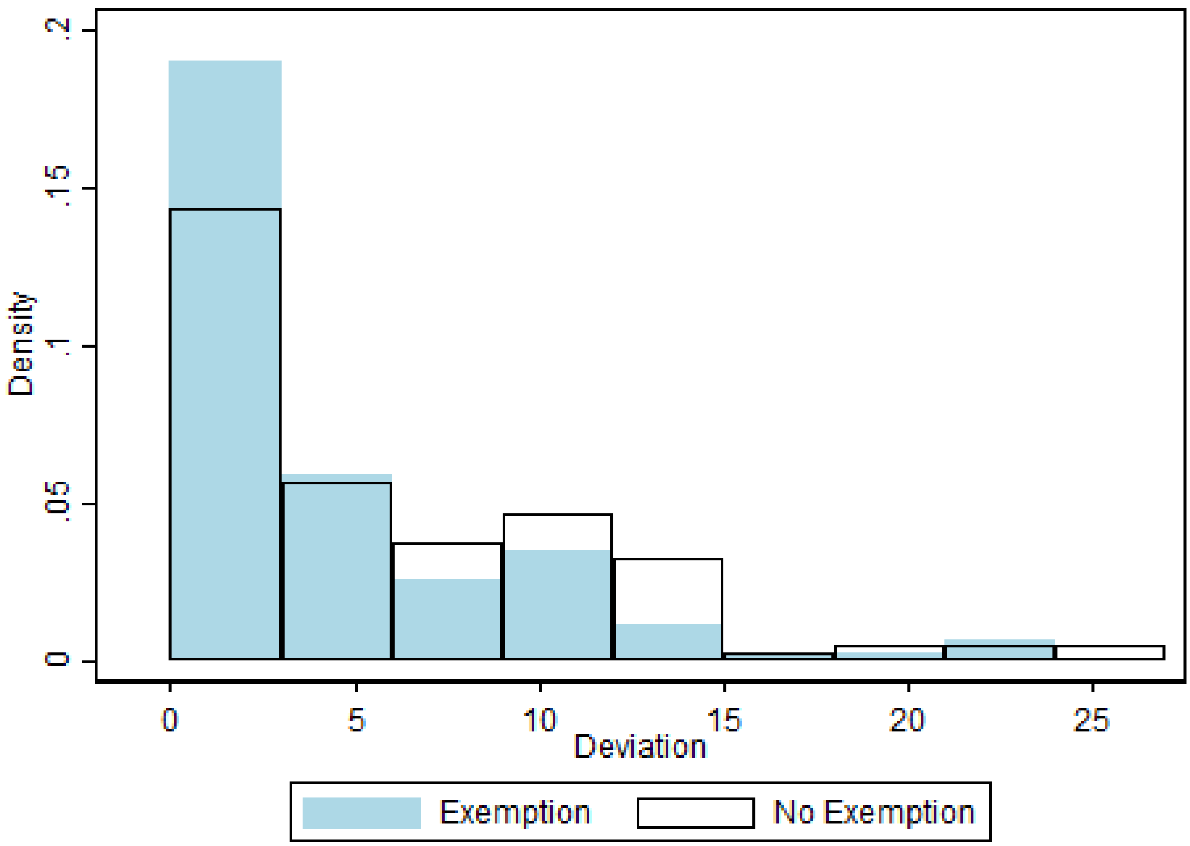


Figure 3.4. Distribution of guesses of the enlarged Hui-Han CEE gap in 2016

Notes: This figure depicts the distribution of the guess accuracy for the treatment $T_{info} * T_{religion}$ (information and exemption; in the blue bins) and the treatment $T_{info} * C_{religion}$ (information but no exemption; the white bins). Each bin covers a 3-point interval. The vertical axis is the density of distribution. The horizontal axis describes how much students' guess is off the accurate information we provide about the CEE gap.

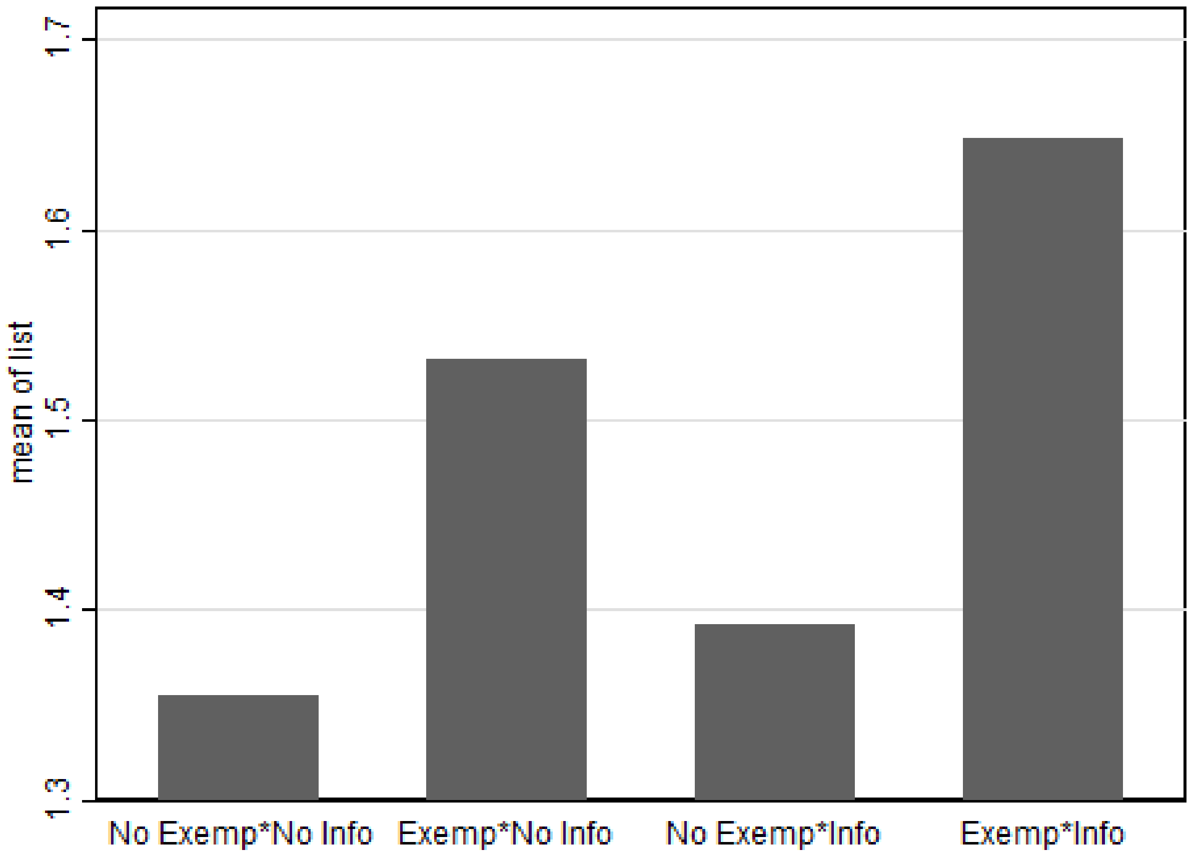


Figure 3.5. Mean agreed response items for the list experiment across treatments

Notes: This graph plots the mean of the number of statements that students agreed with for each treatment in the list experiment. Students were asked "among the following five statements, how many do you agree with?" In the five statements, one of the statements is about fasting attitudes, which we are interested in, and the other four statements concern ways of CEE preparation which are irrelevant to students' religion, with answers depending on students' preferences. Among 5 statements in the list experiment, students can choose to agree with 0-5 of them without specifying which statements exactly they agree with.

3.8 Tables

Table 3.1. Impacts of Ramadan on CEE score

	Outcome = CEE score			
	(1)	(2)	(3)	(4)
Hui	-13.388***	-2.082	-15.021***	-3.655***
	(1.915)	(1.834)	(0.813)	(0.825)
Hui*Year 2012	-2.330	-2.354		
	(2.706)	(2.548)		
Hui*Year 2013	-1.658	-2.212		
	(2.641)	(2.486)		
Hui*Year 2014	-3.530	-2.228		
	(2.603)	(2.451)		
Hui*Year 2015	-0.569	-1.045		
	(2.568)	(2.419)		
Hui*Year 2016	-15.038***	-13.133***		
	(2.556)	(2.407)		
Hui*Ramadan			-13.405***	-11.554***
			(1.877)	(1.767)
STEM-Year FE	Yes	Yes	Yes	Yes
County FE	No	Yes	No	Yes
Gender FE	No	Yes	No	Yes
Mean of Dep. Variable	383.323			
Observations	124,335			
R-squared	0.025	0.136	0.025	0.136

Notes: This table presents the effects of taking the CEE during Ramadan on the relative performance of Muslim students. In columns 1 and 2, we interact Muslim dummy with year dummies, and see an abrupt increase the Hui-Han gap in 2016, the year that Ramadan overlaps with the CEE. In columns 3 and 4, we aggregate the pre-treatment years into a larger control group, and get quantitatively similar results. In columns 1 and 3, we control for STEM-by-Year FE; in columns 2 and 4, we control for County FE and Gender FE. Standard errors in parentheses are clustered at the high school level. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 3.2. 2*2 experimental design

	Information	
Religion	Read the gap in the CEE score between Muslim and Non-Muslim students	Read the gap in the GDP pc between China and Japan
Exemption to delay fast for the CEE	Exemp*Info	Exemp*No Info
No Exemption	No Exemp*Info	No Exemp*No Info

Notes: This table summarizes the 2*2 design of the survey experiment. Randomly, half of the Muslim students receive exemptions to delay fast until after the CEE, while the other half of students do not receive such exemptions. Then the two groups are cross-randomized, such that half of them are required to read a graph on "Hui-Han CEE gap" (Figure 3.1), while the other half of them required to read a graph on "Sino-Japanese income gap" (Figure 3.3).

Table 3.3. Balance tests

	All		No Exp*No Info	Exp*No Info	No Exp*Info	Exp*Info	ANOVA Test	
Variables	Mean	Std.Dev	Mean	Mean	Mean	Mean	F-stat	p-value
Geneder: male	0.405	0.491	0.445	0.398	0.393	0.387	0.38	0.765
Parents with college education	0.045	0.208	0.016	0.047	0.044	0.070	1.57	0.195
Access to computer at home	0.390	0.488	0.390	0.375	0.400	0.394	0.06	0.980
Access to Internet at home	0.814	0.389	0.859	0.758	0.837	0.803	1.67	0.172
Boarding at school	0.831	0.375	0.852	0.82	0.859	0.796	0.84	0.475
Risk loving	2.461	2.125	2.480	2.438	2.652	2.282	0.71	0.548
Perceived value of college	3.692	1.186	3.543	3.680	3.919	3.620	2.51	0.058*
STEM track	0.610	0.488	0.609	0.625	0.630	0.577	0.32	0.810
Honors class	0.334	0.472	0.320	0.336	0.385	0.296	0.88	0.454
Pray everyday	0.589	0.492	0.641	0.555	0.607	0.556	0.95	0.418
Never broke a fast	0.535	0.499	0.602	0.469	0.504	0.563	1.85	0.137
Mock exam score	365.856	62.899	371.006	368.126	366.081	358.953	0.91	0.435
Observations	533		128	128	135	142		

Notes: These two panels present the balance tests across the four different arms in the 2*2 experimental design. Covariates are well-balanced, indicating that the randomization was well-implemented. "Risk loving" and "Perceived value of college" are measured using a five-point Likert scale. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 3.4. The effect of exemption under unawareness

	Estimated CEE gap		
	(1)	(2)	(3)
Exemption*No Info	-0.070	-0.217	-0.148
*(vs. No Exemp*No Info)*	(1.000)	(1.014)	(1.059)
Constant	-17.933***		
	(0.708)		
Class FE	No	Yes	Yes
Covariates	No	No	Yes
Control mean	-17.933	-17.933	-17.933
Observations	247	247	247
R-squared	0.000	0.116	0.218

Notes: This table presents the effects of religious intervention alone on updating prior. As shown in the table, the mean of the elicited 2016 Hui-Han gap is -17.93, close to the -16.4 gap between 2011 and 2015, much smaller than the true value of -29.4, indicating that Muslim students have acute downward bias in their priors. Receiving the exemption to delay the fast does not update this prior in any substantial way. Robust standard errors are in parentheses. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 3.5. The effect of exemption on graph reading (Hui-Han CEE gap)

	Estimated CEE gap from graph reading			Deviation		
	(1)	(2)	(3)	(4)	(5)	(6)
Exemption*Info	-1.903**	-1.988***	-2.199***	-1.644***	-1.663***	-1.862***
*(vs. No Exmp*Info)*	(0.739)	(0.739)	(0.745)	(0.630)	(0.635)	(0.640)
Constant	-24.395***			5.858***		
	(0.529)			(0.451)		
Class FE	No	Yes	Yes	No	Yes	Yes
Covariates	No	No	Yes	No	No	Yes
Control mean	-24.395	-24.395	-24.395	5.858	5.858	5.858
Observations	277	276	274	277	276	274
R-squared	0.024	0.151	0.233	0.024	0.144	0.227

Notes: This table presents the effects of receiving exemption to delay fast on the accuracy of reading the 2016 enlarged Hui-Han gap in CEE performance. The average gap read by students is -24.4, 5 points smaller than the true value of -29.4; receiving an exemption would make the guess 2 points closer to the true value. As shown in columns 4-6, using the “absolute deviation from true value” as the outcome variable produces similar results. Robust standard errors are in parentheses. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 3.6. Extensive margin adjustment

	Accuracy		
	(1)	(2)	(3)
Exemption*Info	0.115*	0.131**	0.157**
*(vs. No Exmp*Info)*	(0.059)	(0.062)	(0.063)
Constant	0.378***		
	(0.043)		
Class FE	No	Yes	Yes
Covariates	No	No	Yes
Control mean	0.378	0.378	0.378
Observations	277	276	274
R-squared	0.013	0.077	0.151

Notes: This table presents the effect of receiving an exemption on the probability of reading the 2016 Hui-Han CEE gap highly accurately. "Accuracy" is a dummy variable indicating whether the deviation from true value (-29.4) is within two points. Students with exemptions are 16 percentage points more likely to make such accurate guesses. Robust standard errors are in parentheses. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 3.7. Heterogeneity based on fasting history

	Estimated CEE gap from graph reading			Deviation		
	(1)	(2)	(3)	(4)	(5)	(6)
Exemption*Info	-0.582	-0.937	-0.865	-0.287	-0.566	-0.582
*(vs. No Exmp*Info)*	(1.067)	(1.073)	(1.094)	(0.909)	(0.924)	(0.937)
Fast (=1)	2.581**	2.889***	2.974***	2.183**	2.391***	2.485***
	(1.043)	(1.052)	(1.075)	(0.888)	(0.905)	(0.922)
Exemption*Fast	-2.618*	-2.201	-2.548*	-2.639**	-2.225*	-2.446*
	(1.462)	(1.486)	(1.535)	(1.245)	(1.279)	(1.315)
Constant	-25.695***			4.758***		
	(1.740)			(0.630)		
Class FE	No	Yes	Yes	No	Yes	Yes
Covariates	No	No	Yes	No	No	Yes
Control mean	-24.395	-24.395	-24.395	5.858	5.858	5.858
Observations	277	276	274	277	276	274
R-squared	0.045	0.177	0.242	0.046	0.167	0.238

Notes: This table presents heterogeneous treatment effects of exemption based on fasting history. Students who strictly followed the Ramadan fasting during high school had larger downward bias to start with, and responded to the religious intervention by eliminating such cognitive bias. On the contrary, students who did not strictly follow Ramadan fasting were not responsive to the exemption. Robust standard errors are in parentheses. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 3.8. Heterogeneity based on valuation for college

	Estimated CEE gap from graph reading			Deviation		
	(1)	(2)	(3)	(4)	(5)	(6)
Exemption*Info	-0.630	-0.828	-1.513	-0.138	-0.271	-0.787
*(vs. No Exmp*Info)*	(1.146)	(1.036)	(1.130)	(1.002)	(0.810)	(0.950)
High Stake (=1)	1.625	1.824*	1.461	1.709*	1.972**	1.558*
	(0.980)	(0.993)	(1.009)	(0.876)	(0.820)	(0.895)
Exemption*High Stake	-1.958	-1.606	-1.053	-2.248*	-1.994*	-1.669
	(1.218)	(1.179)	(1.265)	(1.199)	(1.054)	(1.097)
Constant	-25.462***			4.686***		
	(0.980)			(0.813)		
Class FE	No	Yes	Yes	No	Yes	Yes
Covariates	No	No	Yes	No	No	Yes
Control mean	-25.371	-25.371	-25.371	5.015	5.015	5.015
Observations	274	274	274	274	274	274
R-squared	0.032	0.166	0.232	0.036	0.162	0.230

Notes: This table presents heterogeneous treatment effects of exemption based on perceived valuation of college education. Students who reported higher valuation for college had larger downward bias to start with, and responded to the religious intervention by eliminating such cognitive bias. On the contrary, students who did not value college education were not responsive to the exemption. Robust standard errors are in parentheses. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 3.9. Effect of exemption on graph reading (GDP per capita)

	Estimated GDP gap from graph reading		
	(1)	(2)	(3)
Exemption*No Info	-712.084	-876.285	-1126.323
*(vs. No Exmp*No Info)*	(1088.079)	(1146.520)	(1202.963)
Constant	-28433.923***		
	(760.942)		
Class FE	No	Yes	Yes
Covariates	No	No	Yes
Control mean	-28433.923	-28433.923	-28433.923
Observations	229	229	228
R-squared	0.002	0.061	0.161

Notes: This table presents the placebo effect of receiving an exemption on the accuracy of reading the 2016 Sino-Japanese GDP gap. The religious intervention has no meaningful impacts on reading the GDP gap. Robust standard errors are in parentheses. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 3.10. Active information distortion

		Accuracy	
	(1)	(2)	(3)
Exemption*No Info	-0.031	-0.036	-0.037
	(0.062)	(0.062)	(0.063)
No Exemption*Info	-0.177***	-0.184***	-0.167***
	(0.061)	(0.062)	(0.062)
Exemption*Info	-0.062	-0.057	-0.044
	(0.061)	(0.061)	(0.062)
Constant	0.555***		
	(0.044)		
Class FE	No	Yes	Yes
Covariates	No	No	Yes
Control mean	0.555	0.555	0.555
Observations	533	532	529
R-squared	0.018	0.079	0.126

Notes: This table compares the accuracy of graph-reading across the four arms. Outcome variable is "accuracy," which is defined as deviating within 2 points if reading the Hui-Han CEE gap, or deviating within $1000 when reading the Sino-Japanese income gap. Students are least accurate when asked to read Hui-Han CEE gap without exemptions. Robust standard errors are in parentheses. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 3.11. Fasting attitudes

	Number of Agreed Statements		
	(1)	(2)	(3)
Exemption*No Info	0.177*	0.192*	0.217*
	(0.107)	(0.109)	(0.111)
No Exemption*Info	0.038	0.054	0.049
	(0.105)	(0.107)	(0.109)
Exemption*Info	0.294***	0.299***	0.322***
	(0.104)	(0.106)	(0.108)
Constant	1.354***		
	(0.075)		
Class FE	No	Yes	Yes
Covariates	No	No	Yes
Control mean	1.354	1.354	1.354
Observations	532	531	528
R-squared	0.019	0.053	0.088

Notes: This table presents the effects of the information treatment, the religious treatment, and their interaction on the number of statements one agreed with in the list experiment with in total five statements. The results suggest that receiving the exemption alone makes one more willing to delay fast, receiving the information does not have any significant impact, and receiving both the religious and information interventions has the most powerful persuasion effects. Robust standard errors are in parentheses. * significant at 10%, ** significant at 5%, *** significant at 1%.

3.9 Setup of the Theoretical Model

There are two periods, period 0 and period 1. Student i derives payoff v_i from fasting in each normal Ramadan period. Denote her vulnerability to hunger and thirsty by $\rho_i \in \{0, 1\}$, which she cannot observe directly in period 1. However, she has a prior about this vulnerability which can be fully characterized by $\hat{\rho} \equiv P\{\rho_i = 1\}$.[39] Denote her fasting behavior in periods 0 and 1 by f_0 and f_1 respectively.

Period 0 describes students' fasting behavior during a normal Ramadan period, when Ramadan does not overlap with the CEE. In this period, fasting only affects students' performance in CEE via negatively affecting the effectiveness of learning during Ramadan but not their health status during the exam. The quantity of this effect is expressed as κh, where h is the full effect had she fasted during the CEE and $\kappa < 1$ captures the relatively minor impact on CEE due to inefficient learning during previous Ramadan months. $\omega_i > 0$ represents the importance students i attach to the final outcome of the college entrance exam. For simplicity, we assume that students know their ρ_i due to repeated fasting experience in middle school. They choose f_0 to maximize:

$$f_0 v_i + (1 - f_0)(\kappa \omega_i h \rho_i + \varepsilon_{i0}) \tag{3.5}$$

where ε_{i0} is a random disturbance governed by distribution $F_0(\varepsilon_{i0})$. Put it in another way, students will either fast ($f_0 = 1$), in which case they derive utility v_i by committing to religious practice, or not to fast ($f_0 = 0$), in which case they enjoy enhanced learning effectiveness. Note that we arrange the utility in this form to highlight the tradeoff between fasting ($f_0 = 1$) and not fasting ($f_0 = 0$).

In period 1, students have answered the survey we distributed, and were expecting the CEE in a month. They decide to fast or not in the exam, get anticipatory utility about her

[39] Here we binarize the impact of fasting to be either "negative" or "nonexistent," this is without much loss of generality because no more than 3% of the students in any treatments have beliefs that Ramadan will help boost their performance in the CEE.

performance in the exam and derive utility from fasting behavior, denoted by f_1. In this period, they can no longer remember ρ_i but instead, they form a posterior about ρ_i based on prior $\hat{\rho}$ and previous fasting behavior f_0 as a Bayesian. This is due to either forgetfulness or that they lack knowledge about the impact of fasting on test performance (remember in period 0 that they only experienced fasting when no formal exams like CEE happened). In this period students jointly choose $(\hat{\rho}, f_1)$ to maximize:

$$f_1 \cdot (v_i r - \omega_i E[\rho_i|\hat{\rho}, f_0]h - C(\rho_0 - \hat{\rho})) + (1 - f_1)(-C(\rho_0 - \hat{\rho}) + \varepsilon_{i1}) \tag{3.6}$$

where ε_{i1} is governed by distribution $F_1(\varepsilon_{i1})$. Denote the joint distribution of (ε_{i1}, v_i) and the marginal distribution of v_i by $F(\varepsilon_{i1}, v_i)$ and $G(v_i)$ respectively. Note that v_i has to be non-negative, which is the only restriction for distribution $F(\varepsilon_{i1}, v_i)$ and $G(v_i)$. r is the special return for this special Ramadan period (i.e. fasting during CEE). For simplicity, $r \equiv r_C = 1$ if students regard this fasting period the same and the rest; $r \equiv r_T$ with $0 < r_T < 1$ if students are persuaded by religious leaders, and believe that fasting may not be necessary during the particular exam days. Therefore $v_i r$ captures the payoff from fasting during CEE. $-E[\rho_i|\hat{\rho}, f_0]h$ is the expected cost of fasting during CEE, and $-C(\rho_0 - \hat{\rho})$ is the cognitive cost of manipulating her prior away from her original prior ρ_0 had motivated beliefs been not at play. We assume that $C(.)$ is twice continuously differentiable, minimized at 0. We also assume ρ_0 and $\hat{\rho}$ to be a real number between 0 and 1. Note that we arrange the utility in this form to highlight the utility derived from both fasting ($f_1 = 1$) and not fasting ($f_1 = 0$) respectively.

The major difference between our model and the previous studies is the focus on the manipulable prior $\hat{\rho}$, which merits further discussions. Aside from the mechanical explanation above, another interpretation of ρ_0 is that this prior is subconscious, and the subject's cognition process manipulates her prior away from the subconscious one to maximize her anticipated utility. The modeling of $\hat{\rho}$ is similar in spirit to Augenblick *et al.* (2016), where students manipulate their beliefs about the probability of dooms day above their original beliefs had a religious concern not been present. Importantly, this subconscious belief need not be accurate. While Augenblick et al. are agnostic about the formation and implications of differential ρ_0 in their paper as this is not their focus, we directly test the additional implication of a wrong ρ_0 and confirms the validity of our model.

Our model is also different from previous studies on motivated beliefs in that the anticipatory utility merely comes from students' expectation about their own performance in the exam. Arguably, as an once-in-lifetime high-stake exam, for which students have been preparing for years, the effect of anticipatory utility should be particularly strong. We do not specifically model the utility of religious beliefs, such as utility carried by h itself, which may reflect people's belief on how omnipotent their religion is. The primary reason of this omission is that the incorporation of this utility does not qualitatively change our results, and our empirical results do not support this possibility either.

This model has a number of predictions about students' response in beliefs and fasting attitudes. We categorize them into three groups to highlight the relationship between these propositions and the results presented in the next section. Specifically, Proposition 1 predicts response in beliefs under unawareness of fasting impact; Proposition 2, 3, 4 predicts response in beliefs under awareness of fasting impact; Proposition 5 discusses the relationship between the beliefs and the perceived importance of the College Entrance Exam; Proposition 6 and 7 presents our model's prediction on fasting attitudes.

Proposition 1 When $\rho_0 = 0$, $\hat{\rho} = 0$ irrespective of the value of f_0, f_1, r and v_i.

This proposition discusses how students might react when their subconscious beliefs are wrong. Since anecdotal evidence suggests that students may not be aware of the negative impact of fasting at all, our proposition focus on the prediction in this case. The framework predicts that the students do not have to incur any cost to create illusion, but just happily take the view that fasting does not do even cause the slightest harm. As a result they sincerely do not believe that on average, fasting is significantly detrimental to their cognitive function regardless of whether religious leader try to persuade them to fast or not during

CEE. This prediction of this proposition, in our context, is elaborated by Hypothesis 1 in the main text.

Proposition 2 In case of $\rho_0 > 0$ and for almost any given (ε_{i1}, f_0), $\hat{\rho} < \rho_0$ if $f_1 = 1$ for any positive r and v_i.

As one of the most basic results of this model, this proposition says that for people who choose to fast, they have the incentive to distort their prior as long as they become partially aware of the fact that fasting is harmful to their exam performance, irrespective of its magnitude. In our experiment, we use "belief about the average impact of taking the CEE during Ramadan" as a proxy for the parameter ρ_0. This prediction of this proposition, in our context, is elaborated by Hypothesis 2 in the main text.

As r change, students have different incentives to fast, hence to distort their beliefs about the impact of fasting, as illustrated in the following proposition:

Proposition 3 In case of $\rho_0 > 0$, for any given (ε_{i1}, f_0), $\hat{\rho}$ is weakly decreasing as r increases: for small r such that students opt not to fast, their belief $\hat{\rho} = \rho_0$. Students do not bias belief downwards until r is large enough such that $f_1 = 1$, and the extent of distortion is constant for $f_1 = 1$.

Put it in another way, this proposition says that given the awareness of the negative impact of fasting on exam, students are more likely to form the right belief and not to fast if the religious leader successfully persuade them not to do so by lowering their r. Moreover, as students face two options for f_1 (to fast or not), their choice regarding the extent of manipulation is also binarized: either they do not manipulate at all and do not fast, or they manipulate to a constant extent and and stick to the religious practice regardless of how valuable/harmful it is. Students' adjustment on their prior is therefore entirely on the extensive margin. This prediction of this proposition, in our context, is elaborated by Hypothesis 3 and 4 in the main text.

For simplicity, we additionally assume that κ is small in the discussion of last proposition. This assumption says that the impact of Ramadan fasting during pre-exam period (say fasting one year or two years ahead of the CEE) is minor to fasting on CEE exam. We argue that this is a reasonable assumption for the following two reasons: first, the length of Ramadan fasting is merely one month for every year in Islamic calendar, which is relatively short compared to years of exam preparation; second, even if students' learning activity are affected during fasting, they can still make up for it by studying harder before/after the fasting month.

This proposition concerns the heterogeneity of the treatment effects with respective to past fasting behavior f_0. While the prediction of model in general may not be entirely clear given different κ, f_0, $C(.)$ and the joint distribution of v_i and ρ_i, with assumption on κ, we can derive the following results.

Proposition 4 When κ is sufficiently small, the distribution of v_i given $f_0 = 1$ stochastically dominates that given $f_0 = 0$. Hence given the same ρ_0, $E[\hat{\rho}|f_0 = 1] < E[\hat{\rho}|f_0 = 0]$

This proposition discuss the case where fasting in the past can barely affect the CEE outcome. In this case, students can only extract information about v_i from f_0. For those who did not fast in the past, then have lower v_i, hence less incentive to manipulate their beliefs. We view this assumption as plausible because as we have discussed in institutional details, past fasting rarely affects the exam outcome because students have three years to prepare for the exam, hence they can have plenty time and opportunities to make up had they, by any chance, fallen behind during the fasting period. Moreover, the results are fairly robust even when κ is large [40] This prediction of this proposition, in our context, is elaborated by Hypothesis 5 in the main text.

Proposition 5 Holding other parameters constant and $f_1 = 1$, $\hat{\rho}$ is weakly decreasing as ω_i increases.

The intuition of this proposition is clear: the motivation of distortion is determined by the value of anticipatory utility. If the individual in question attaches more importance to

[40] When κ is large, f_0 is affected by both ρ_i and v_i. We need to consider the joint distribution of these two variables. However, even in this case, with moderate assumptions on cognitive cost, we will be able to get the result that the optimal probabilistic beliefs for those who choose f_1 to be 1 is smaller for students who fast in the past.

the exam, she naturally cares more about the value of her anticipatory utility. Hence, given that she decide to fast, the presence of motivated beliefs generates more biases when the stake is even higher, which runs against the argument that economic importance may mitigates distortion in this particular case. This prediction of this proposition, in our context, is elaborated by Hypothesis 6 in the main text.

The last two propositions concerns treatment effects on fasting attitudes.

Proposition 6 In case of $\rho_0 > 0$, for almost any given $(\varepsilon_{i1}, f_0, v_i, r)$, as long as $h > 0$, $f_1 = 0$ if and only if $\hat{\rho} = \rho_0$

This proposition provides us with a tight link between the elicited beliefs $\hat{\rho}$ and fasting behavior f_1 during CEE: when students are aware of the harm of Ramadan fasting (i.e. their subconscious belief ρ_0 is positive), those who hold the right beliefs will not fast and vice versa. While the implication that we can precisely identify those who do not fast must express the right belief is not robust to alteration such as incorporating people's utility from the omnipotence of their religion (i.e. utility as a function of h), it is indeed robust that given a correct ρ_0, as beliefs become more accurate, students are less likely to fast during the CEE across different treatment groups. This proposition provides a way to proxy fasting behavior: if we want to focus on the group of people who fast (say, examine the impact of perceived stake on biases conditional on fasting), we can restrict our attention to subsample where people don't read the graph accurately.

While people will not adjust their beliefs given the initial unawareness of the harm of fasting, the persuasion from religious leaders do decrease r, which decreases the gap of utility between fasting and not fasting in period 1. If there are any independent disturbance of fasting preferences as illustrated by ε_{i1} in the model, the rate of fasting will also be decreased by authorization from religious leaders.

The next proposition discusses the effectiveness of information treatments in terms of changing fasting attitudes. We can easily deduce from Equation 3.6 that religious leader persuasion alone is sufficient to shift the fasting decisions of some people. In addition to that direct channel, there is also an additional role of information dissemination on changing fasting attitudes:

Proposition 7 For any given ε_{i1}, Denote the minimum level of v_i needed to choose fast for treatments "*No Exemp*No Info*," "*Exemp*No Info*," "*No Exemp*Info*," "*Exemp*Info*" by $\bar{v}_1$, $\bar{v}_2$, $\bar{v}_3$, $\bar{v}_4$, respectively. If, say, any non-negative v_i is enough for fast in treatment "*No Exemp*No Info*," then $\bar{v}_1 = 0$. We have: (i) $\bar{v}_1 < \bar{v}_2$, $\bar{v}_1 < \bar{v}_3$;(ii) $\bar{v}_4 - \bar{v}_2 > \bar{v}_3 - \bar{v}_1$.

This proposition use a specific set measures, $\bar{v}_1$, $\bar{v}_2$, $\bar{v}_3$, $\bar{v}_4$, to measure people's preference to choose fasting in the end. The higher the threshold is, to the less extent people would prefer fasting. (i) says that the threshold for merely providing information $\bar{v}_3$ and threshold for merely providing religious exemption $\bar{v}_2$ both move up relative to control threshold $\bar{v}_1$, indicating that both treatment works in the same direction, whereas the relative effectiveness of them is an empirical question. (ii) says that the information treatment and religious exemption may serve as compliments: when religious exemption is granted, the effectiveness of providing information in terms of the movement of the threshold, $\bar{v}_4 - \bar{v}_2$, is larger than $\bar{v}_3 - \bar{v}_1$, in which case no exemption is granted. Of course, the results still hold when we regard these threshold as a function of ε_{i1}, and integrate over it to compare the expected level of thresholds. This prediction of this proposition, in our context, is elaborated by Hypothesis 8 in the main text.

3.10 Reading Materials

3.10.1 Reading on Exemption - The Treatment Condition

Between 2016 and 2018, the Muslim holy month of Ramadan coincided with the college entrance examination. Therefore, for many Muslim students, "whether they can break the fast and make it up later after the college entrance exam" has become an important issue that cannot be ignored.

In order to understand whether "Ramadan fasting can be postponed during the college entrance examination," we consulted Guo Haihui, a well-known scholar who graduated from the Royal Religious University of Malaysia and the current Imam of the century-old temple "Xiangfang Mosque." He said:

"The acts of worship of Islam has three goals: to express faith to Allah, exercise good words and deeds and sublimate souls. The Prophet (PBUH) said: 'Allah does not look at your appearance and your goods. He looks only at your heart and your deeds.' The good intention for any deed is the key to get good results. The college entrance examination has become a major concern for the whole society, let alone for the students. It is no exaggeration to describe it as the turning point for the students. Because the examination is both mentally and physically exhausting and no easier than any other work, both parents and students need to make great efforts to prepare for it. Therefore, it is necessary to appropriately reduce their burden. To temporarily postpone the fasting during the college entrance examination will neither anger Allah, nor will it weaken your beliefs."

We also consulted the famous scholar Liu Xueqiang, who is also the vice president of the Provincial Islamic Association and the Imam of the famous Xigong Mosque. His suggestion was consistent with that of Guo Haihui:

"The purpose of Islamic law is to create convenience for people, not to create difficulties. The implementation of Islamic law can be flexible in the actual process and it should not be

interpreted rigidly. Allah never asks people to do things beyond their ability. Therefore, if the candidate thinks that fasting will affect his or her test scores, it is acceptable to break the fast, and make up afterwards. It poses no problem in the Islamic law."

This situation is not unique to China: as the college entrance examination is held in June in many countries, the jurists in these countries also give corresponding doctrinal orders for the examination and fasting. Through summarizing, we find that many authoritative religious scholars and institutions abroad share similar views on this issue with imams in China. For example, when being asked if "students can break the fast during the college entrance examination," Grand Mufti Shawki Allam of the Egyptian Shariah Committee replied:

"If fasting affects the students' ability to revise and study for the exam, resulting in symptoms like reduced concentration, unresponsiveness, dizziness, etc., and the exam time stipulated by the education system cannot be adjusted to the end of Ramadan, students should break the fast and make it up after the exam, so that their previous efforts will not be wasted."

Experts of the French Muslim Religious Committee also conducted in-depth researches on this issue and finally issued a notice: "It is recommended that candidates break the fast, especially those who need to take the exam in the afternoon. However, they need to make it up after Ramadan."

3.10.2 Reading on Art - The Control Condition

There is a US diplomat who spent ten years in Moscow in the 1920s and 1930s. He wrote in his memoir that he has watched the "Swan Lake" performance for 300 times. Even for a classic ballet as famous as the "Swan Lake," 300 times is too much. But for a diplomat, some social engagements are inevitable, and he had no choice but to watch this play again and again until it was a bit overwhelming.

I guess, for the first few dozen times to watch the "Swan Lake" performance, what the American heard was the beautiful music of Tchaikovsky and what he saw was the beautiful performance of the artists of the former Soviet Union. He appreciated it wholeheartedly and *applauded ardently from time to time. After having watched it for 100 times, the impression became different. At that time, he could only hear some instruments ringing and see some people running on the stage and he became slow-witted as well. Then, after 200 times, the impression changed again. The music was on and the curtain was up, but there was only the white void in front of him - he was caught in the nightmare of this play. At this point, his eyes were blank, his face was smirking, like a hibernating crocodile whose loose muscles could not support the chin, or a landing boat rushing to the beach, and his mouth was opening, with big drops rolling down from the corner of his mouth and falling on his knees. It was so intoxicating that not until the curtain was down and someone switched off the light did he realize that it was over. He quickly slapped himself awake and went home. Later, when he got the order to leave the Soviet Union, he said with relief: well, finally, no more "Swan Lake."*

As you know, the scene above is just my guess - to be honest, no one will ever include this in one's memoirs - but I think anyone repeatedly appreciating a piece of work will encounter these three phases. In the first phase, you hear the music and see the dance - in short, you are enjoying art. In the second phase, you hear some sounds and see some objects moving, and you are aware of a familiar physical process. In the third phase, you have gained a philosophical perspective and finally realized that the ballet, just like everything else in the world, is a form of material existence. From art to science and then to philosophy, it is a process of returning to the original nature.

Normally, people's appreciation always stays in the first phase, but some people can reach the second phase. For example, in the movie "Farewell My Concubine," the tyrant played by Ge You blamed an actor: the Conqueror played by other people took six steps, why did you take four steps? In the lab, a physicist would also ask an object in confusion: how can your acceleration be two Gs while others is a G when falling in a vacuum? In the laboratory, a physical process must be reproducible, or otherwise it will not be scientific. Therefore, no object falls with two Gs' acceleration. The classic works of art should also be reproducible.

Take "Swan Lake" for example, the content of this ballet cannot be changed in order to let future generations appreciate the best things created by the predecessors. It can only be played over and over again.

Classic works are good and worth watching, but not too many times. Otherwise, the art cannot be appreciated - just like tea drinking in the "Dream of Red Mansions": one cup is for tasting, two cups are for the thirst, and three cups are drinking like a fish. Of course, whether it is tea-tasting or drinking like a fish, it is just a way of material existence. In this respect, there is no difference between them...

BIBLIOGRAPHY

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

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

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

Afifi, ZEM, (1997). "Daily practices, study performance and health during the ramadan fast," *Journal of the Royal Society of Health*, *117*(4), pp. 231–235.

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

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

Almond, Douglas and Bhashkar Mazumder, (2011). "Health capital and the prenatal environment: The effect of Ramadan observance during pregnancy," *American Economic Journal: Applied Economics*, *3*(4), pp. 56–85.

Almond, Douglas, Bhashkar Mazumder, and Reyn Van Ewijk, (2015). "In utero Ramadan exposure and children's academic performance," *The Economic Journal*, *125*(589), pp. 1501–1533.

Ambuehl, Sandro, (2017). "An offer you can't refuse? Incentives change how we inform ourselves and what we believe,"Technical report, CESifo Group Munich.

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

Athey, Susan and Guido W Imbens, (2017). "The econometrics of randomized experimentsa," In *Handbook of Economic Field Experiments*, *1*: Elsevier, pp. 73–140.

Augenblick, Ned, Jesse M Cunha, Ernesto Dal Bó, and Justin M Rao, (2016). "The economics of faith: Using an apocalyptic prophecy to elicit religious beliefs in the field," *Journal of Public Economics*, *141*, pp. 38–49.

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

Avery, Christopher and Caroline Minter Hoxby, (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*: University of Chicago Press, pp. 239–302.

AZZI, CORRY AND RONALD EHRENBERG, (1975). "Household allocation of time and church attendance," *Journal of Political Economy*, *83*(1), pp. 27–56.

BAILEY, MARTHA J AND SUSAN M DYNARSKI, (2011). "Inequality in postsecondary education," In G.J. Duncan and R.J. Murnane (eds.) *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*: New York: Russell Sage, pp. 117–132.

BALINSKI, MICHEL AND TAYFUN SÖNMEZ, (1999). "A tale of two mechanisms: Student placement," *Journal of Economic Theory*, *84*(1), pp. 73–94.

BARLEVY, GADI AND DEREK NEAL, (2012). "Pay for percentile," *American Economic Review*, *102*(5), pp. 1805–31.

BARRERA-OSORIO, FELIPE AND DHUSHYANTH RAJU, (2017). "Teacher performance pay: Experimental evidence from Pakistan," *Journal of Public Economics*, *148*, pp. 75–91.

BELASCO, ANDREW S, (2013). "Creating college opportunity: School counselors and their influence on postsecondary enrollment," *Research in Higher Education*, *54*(7), pp. 781–804.

BÉNABOU, ROLAND, (2015). "The economics of motivated beliefs," *Revue d'économie politique*, *125*(5), pp. 665–685.

BÉNABOU, ROLAND AND JEAN TIROLE, (2011). "Identity, morals, and taboos: Beliefs as assets," *The Quarterly Journal of Economics*, *126*(2), pp. 805–855.

BÉNABOU, ROLAND AND JEAN TIROLE, (2016). "Mindful economics: The production, consumption, and value of beliefs," *Journal of Economic Perspectives*, *30*(3), pp. 141–64.

BENARTZI, SHLOMO AND RICHARD THALER, (2007). "Heuristics and biases in retirement savings behavior," *Journal of Economic Perspectives*, *21*(3), pp. 81–104.

BERGMAN, PETER, (2015). "Parent-child information frictions and human capital investment: Evidence from a field experiment," *Working paper*.

BERGMAN, PETER, JEFFREY T DENNING, AND DAYANAND MANOLI, (2019). "Is information enough? The effect of information about education tax benefits on student outcomes," *Journal of Policy Analysis and Management*, *38*(3), pp. 706–731.

BERMAN, ELI, (2000). "Sect, subsidy, and sacrifice: An economist's view of ultra-orthodox Jews," *The Quarterly Journal of Economics*, *115*(3), pp. 905–953.

BERTRAND, MARIANNE, SENDHIL MULLAINATHAN, AND ELDAR SHAFIR, (2004). "A behavioral-economics view of poverty," *American Economic Review*, *94*(2), pp. 419–423.

BETTINGER, ERIC P AND BRENT J EVANS, (2019). "College guidance for all: A randomized experiment in pre-college advising," *Journal of Policy Analysis and Management*, *38*(3), pp. 579–599.

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

Bird, Kelli A, Benjamin L Castleman, Joshua Goodman, and Cait Lamberton, (2017). "Nudging at a national scale: Experimental evidence from a FAFSA completion campaign," *EdPolicy Works Working Paper Series No. 54*.

Bond, Timothy N, George Bulman, Xiaoxiao Li, and Jonathan Smith, (2018). "Updating human capital decisions: Evidence from SAT score shocks and college applications," *Journal of Labor Economics*, *36*(3), pp. 807–839.

Bordon, Paola and Chao Fu, (2015). "College-major choice to college-then-major choice," *The Review of Economic Studies*, *82*(4), pp. 1247–1288.

Bowen, William G, Matthew M Chingos, and Michael S McPherson, (2009). *Crossing the Finish Line: Completing College at America's Public Universities*: Princeton University Press.

Bowman, Nicholas A, Sanga Kim, Laura Ingleby, David C Ford, and Christina Sibaouih, (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*, p. 0162373718778133.

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

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

Burkhardt, JC, SL DesJardins, CA Teener, SE Gay, and SA Santen, (2018). "Predicting medical school enrollment behavior: Comparing an enrollment management model to expert human judgment.," *Academic Medicine: Journal of the Association of American Medical Colleges*, *93*(11S), pp. S68–S73.

Calsamglia, Caterina, Chao Fu, and Maia Güell, (2018). "Structural estimation of a model of school choices: The Boston mechanism vs. its alternatives,"Technical report, National Bureau of Economic Research.

Campante, Filipe and David Yanagizawa-Drott, (2015). "Does religion affect economic growth and happiness? Evidence from Ramadan," *The Quarterly Journal of Economics*, *130*(2), pp. 615–658.

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

CARRUTHERS, CELESTE K AND WILLIAM F FOX, (2016). "Aid for all: College coaching, financial aid, and post-secondary persistence in Tennessee," *Economics of Education review*, *51*, pp. 97–112.

CASTLEMAN, BEN, (2017). "Behavioral insights for federal higher education policy.," *Urban Institute*.

CASTLEMAN, BENJAMIN AND JOSHUA GOODMAN, (2018). "Intensive college counseling and the enrollment and persistence of low-income students," *Education Finance and Policy*, *13*(1), pp. 19–41.

CASTLEMAN, BENJAMIN L, LAURA OWEN, AND LINDSAY C PAGE, (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*, pp. 168–179.

CASTLEMAN, BENJAMIN L AND LINDSAY C PAGE, (2015). "Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates?" *Journal of Economic Behavior & Organization*, *115*, pp. 144–160.

CASTLEMAN, BENJAMIN L AND LINDSAY C PAGE, (2017). "Parental influences on postsecondary aking: Evidence from a text messaging experiment," *Educational Evaluation and Policy Analysis*, *39*(2), pp. 361–377.

CASTLEMAN, BENJAMIN L, SAUL SCHWARTZ, AND SANDY BAUM, (2015b). *Decision making for student success: Behavioral insights to improve college access and persistence*: Routledge.

CHADE, HECTOR, GREGORY LEWIS, AND LONES SMITH, (2014). "Student portfolios and the college admissions problem," *Review of Economic Studies*, *81*(3), pp. 971–1002.

CHALFIN, AARON, OREN DANIELI, ANDREW HILLIS, ZUBIN JELVEH, MICHAEL LUCA, JENS LUDWIG, AND SENDHIL MULLAINATHAN, (2016). "Productivity and selection of human capital with machine learning," *American Economic Review*, *106*(5), pp. 124–27.

CHEN, TING, JAMES KAI-SING KUNG, AND CHICHENG MA, (2017). "Long live Keju! The persistent effects of china's imperial examination system," *Working paper*.

CHEN, YAN AND ONUR KESTEN, (2017). "Chinese college admissions and school choice reforms: A theoretical analysis," *Journal of Political Economy*, *125*(1), pp. 99–139.

CHENG, ING-HAW, SAHIL RAINA, AND WEI XIONG, (2014). "Wall Street and the housing bubble," *American Economic Review*, *104*(9), pp. 2797–2829.

CHETTY, RAJ, (2015). "Behavioral economics and public policy: A pragmatic perspective," *American Economic Review*, *105*(5), pp. 1–33.

COLEMAN, JAMES S, (1988). "Social capital in the creation of human capital," *American Journal of Sociology*, *94*, pp. S95–S120.

CONLISK, JOHN, (1996). "Why bounded rationality?" *Journal of economic literature*, *34*(2), pp. 669–700.

CORCORAN, SEAN P, JENNIFER L JENNINGS, SARAH R COHODES, AND CAROLYN SATTIN-BAJAJ, (2018). "Leveling the playing field for high school choice: Results from a field experiment of informational interventions," *NBER Working Paper No. 24471*.

CURRIE, JANET, (2004). "The take up of social benefits," *NBER Working Paper No. 10488*.

DAMGAARD, METTE TRIER AND HELENA SKYT NIELSEN, (2018). "Nudging in education," *Economics of Education Review*, *64*, pp. 313–342.

DANIEL, BEN, (2015). "Big data and analytics in higher education: Opportunities and challenges," *British Journal of Educational Technology*, *46*(5), pp. 904–920.

DAVIS, JONATHAN, JONATHAN GURYAN, KELLY HALLBERG, AND JENS LUDWIG, (2017). "The economics of scale-up," *NBER Working Paper No. W23925*.

DE REE, JOPPE JAITZE, KARTHIK MURALIDHARAN, MENNO PRADHAN, AND HALSEY ROGERS, (2018). "Double for nothing? Experimental evidence on an unconditional teacher salary increase in Indonesia," *The Quarterly Journal of Economics*, *133*(2), pp. 993–1039.

DEE, THOMAS S AND JAMES WYCKOFF, (2015). "Incentives, selection, and teacher performance: Evidence from impact," *Journal of Policy Analysis and Management*, *34*(2), pp. 267–297.

DEMING, DAVID J AND DAVID FIGLIO, (2016). "Accountability in Us education: Applying lessons from K-12 experience to higher education," *Journal of Economic Perspectives*, *30*(3), pp. 33–56.

DESJARDINS, STEPHEN L, DENNIS A AHLBURG, AND BRIAN P MCCALL, (2006). "An integrated model of application, admission, enrollment, and financial aid," *The Journal of Higher Education*, *77*(3), pp. 381–429.

DESJARDINS, STEPHEN L AND ROBERT K TOUTKOUSHIAN, (2005). "Are students really rational? The development of rational thought and its application to student choice," In *Higher Education: Handbook of Theory and Research*: Springer, pp. 191–240.

DETTLING, LISA J, SARENA GOODMAN, AND JONATHAN SMITH, (2018). "Every little bit counts: The impact of high-speed internet on the transition to college," *Review of Economics and Statistics*, *100*(2), pp. 260–273.

DI TELLA, RAFAEL, RICARDO PEREZ-TRUGLIA, ANDRES BABINO, AND MARIANO SIGMAN, (2015). "Conveniently upset: Avoiding altruism by distorting beliefs about others' altruism," *American Economic Review*, *105*(11), pp. 3416–42.

DILLON, ELEANOR WISKE AND JEFFREY ANDREW SMITH, (2017a). "The consequences of academic match between students and colleges," *Working paper*.

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

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

Dynarski, Susan, CJ Libassi, Katherine Michelmore, and Stephanie Owen, (2018). "Closing the gap: The effect of a targeted, tuition-free promise on college choices of high-achieving, low-income students," *Working Paper*.

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

Eil, David and Justin M Rao, (2011). "The good news-bad news effect: Asymmetric processing of objective information about yourself," *American Economic Journal: Microeconomics*, *3*(2), pp. 114–38.

Exley, Christine and Judd B Kessler, (2018). "Motivated errors," *Working paper*.

Figlio, David N and Joshua Winicki, (2005). "Food for thought: The effects of school accountability plans on school nutrition," *Journal of Public Economics*, *89*(2-3), pp. 381–394.

Finan, Frederico, Benjamin A Olken, and Rohini Pande, (2017). "The personnel economics of the developing state," In *Handbook of Economic Field Experiments*, *2*: Elsevier, pp. 467–514.

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

Fryer, Roland G, (2013). "Teacher incentives and student achievement: Evidence from New York City public schools," *Journal of Labor Economics*, *31*(2), pp. 373–407.

Fryer, Roland G, (2017). "The production of human capital in developed countries: Evidence from 196 randomized field experiments," In *Handbook of Economic Field Experiments*, *2*: Elsevier, pp. 95–322.

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

Gilligan, Daniel O, Naureen Karachiwalla, Ibrahim Kasirye, Adrienne Lucas, and Derek A Neal, (2018). "Educator incentives and educational triage in rural primary schools," *NBER Working Paper No. 24911*.

Glewwe, Paul, Nauman Ilias, and Michael Kremer, (2010). "Teacher incentives," *American Economic Journal: Applied Economics*, *2*(3), pp. 205–27.

GLEWWE, PAUL AND KARTHIK MURALIDHARAN, (2016). "Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications," In *Handbook of the Economics of Education*, *5*: Elsevier, pp. 653–743.

GONZÁLEZ, JULIE M BYERS AND STEPHEN L DESJARDINS, (2002). "Artificial neural networks: A new approach to predicting application behavior," *Research in Higher Education*, *43*(2), pp. 235–258.

GOODMAN, JOSHUA, ODED GURANTZ, AND JONATHAN SMITH, (2018). "Take two! SAT retaking and college enrollment gaps," *NBER Working Paper No. 24945*.

GOODMAN, JOSHUA AND MELANIE RUCINSKI, (2018). "Increasing diversity in Boston's Exam Schools," *Rappaport Institute for Greater Boston Policy Brief*.

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

GOODMAN, SARENA F AND LESLEY J TURNER, (2013). "The design of teacher incentive pay and educational outcomes: Evidence from the New York City bonus program," *Journal of Labor Economics*, *31*(2), pp. 409–420.

HAFALIR, ISA E, RUSTAMDJAN HAKIMOV, DOROTHEA KÜBLER, AND MORIMITSU KURINO, (2018). "College admissions with entrance exams: Centralized versus decentralized," *Journal of Economic Theory*, *176*, pp. 886–934.

HALLAK, MOHAMAD H AND MOHAMMAD ZAFAR A NOMANI, (1988). "Body weight loss and changes in blood lipid levels in normal men on hypocaloric diets during ramadan fasting," *The American Journal of Clinical Nutrition*, *48*(5), pp. 1197–1210.

HARRISON, GLENN W AND JOHN A LIST, (2004). "Field experiments," *Journal of Economic literature*, *42*(4), pp. 1009–1055.

HART, P SOL AND ERIK C NISBET, (2012). "Boomerang effects in science communication: How motivated reasoning and identity cues amplify opinion polarization about climate mitigation policies," *Communication Research*, *39*(6), pp. 701–723.

HASTINGS, JUSTINE, CHRISTOPHER A NEILSON, AND SETH D ZIMMERMAN, (2018). "The effects of earnings disclosure on college enrollment decisions," *NBER Working Paper No. 21300*.

HAYEK, FRIEDRICH AUGUST, (1945). "The use of knowledge in society," *American Economic Review*, *35*(4), pp. 519–530.

HELLER, SARA B, ANUJ K SHAH, JONATHAN GURYAN, JENS LUDWIG, SENDHIL MULLAINATHAN, AND HAROLD A POLLACK, (2017). "Thinking, fast and slow? Some field experiments to reduce crime and dropout in Chicago," *The Quarterly Journal of Economics*, *132*(1), pp. 1–54.

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

HERZOG, SERGE, (2006). "Estimating student retention and degree-completion time: Decision trees and neural networks vis-à-vis regression," *New Directions for Institutional Research*, *2006*(131), pp. 17–33.

HESS, SIMON, (2017). "Randomization inference with stata: A guide and software," *Stata Journal*, *17*(3), pp. 630–651.

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

HOLSINGER, DONALD B AND W JAMES JACOB, (2009). *Inequality in education: Comparative and international perspectives*, *24*: Springer Science & Business Media.

HOSSLER, DON AND KAREN S GALLAGHER, (1987). "Studying student college choice: A three-phase model and the implications for policymakers.," *College and University*, *62*(3), pp. 207–21.

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

HOWELL, JESSICA S AND MATEA PENDER, (2016). "The costs and benefits of enrolling in an academically matched college," *Economics of Education Review*, *51*, pp. 152–168.

HOXBY, CAROLINE AND CHRISTOPHER AVERY, (2013). "The missing "one-offs": The hidden supply of high-achieving, low-income students," *Brookings Papers on Economic Activity*, *2013*(1), pp. 1–65.

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

HOXBY, CAROLINE M AND SARAH TURNER, (2013). "Expanding college opportunities for high-achieving, low-income students," *SIEPR Discussion Paper No. 12-014*.

HURWITZ, MICHAEL, PREEYA P MBEKEANI, MARGARET M NIPSON, AND LINDSAY C PAGE, (2017). "Surprising ripple effects: How changing the SAT score-sending policy for low-income students impacts college access and success," *Educational Evaluation and Policy Analysis*, *39*(1), pp. 77–103.

HURWITZ, MICHAEL, JONATHAN SMITH, SUNNY NIU, AND JESSICA HOWELL, (2015). "The Maine question: How is 4-year college enrollment affected by mandatory college entrance exams?" *Educational Evaluation and Policy Analysis*, *37*(1), pp. 138–159.

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

Hyman, Joshua, (2019). "Can light-touch college-going interventions make a difference? Evidence from a statewide experiment in Michigan," *Journal of Policy Analysis and Management*.

Iannaccone, Laurence R, (1992). "Sacrifice and stigma: Reducing free-riding in cults, communes, and other collectives," *Journal of Political Economy*, *100*(2), pp. 271–291.

Iannaccone, Laurence R, (1998). "Introduction to the economics of religion," *Journal of Economic Literature*, *36*(3), pp. 1465–1495.

Imberman, Scott A and Michael F Lovenheim, (2015). "Incentive strength and teacher productivity: Evidence from a group-based teacher incentive pay system," *Review of Economics and Statistics*, *97*(2), pp. 364–386.

Iyer, Sriya, (2016). "The new economics of religion," *Journal of Economic Literature*, *54*(2), pp. 395–441.

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

"Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools, author=Jacob, Brian, journal=Journal of Public Economics, volume=89, number=5-6, pages=761–796, year=2005, publisher=Elsevier."

Jacob, Brian, (2017). "When evidence is not enough: Findings from a randomized evaluation of evidence-based literacy instruction (EBLI)," *Labour Economics*, *45*, pp. 5–16.

Jacob, Brian and Steven D Levitt, (2003). "Rotten apples: An investigation of the prevalence and predictors of teacher cheating," *The Quarterly Journal of Economics*, *118*(3), pp. 843–877.

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

Jia, Ruixue and Hongbin Li, (2016). "Access to elite education, wage premium, and social mobility: The truth and illusion of china's college entrance exam," *Stanford Center for International Development, Working Paper*(577).

Kahneman, Daniel, (2003). "Maps of bounded rationality: Psychology for behavioral economics," *American Economic Review*, *93*(5), pp. 1449–1475.

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

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

Kane, Thomas J, (1999). *The Price of Admission: Rethinking How Americans Pay for College*: Brookings Institution Press.

Kang, Chungseo and Darlene Garcia Torres, (2018). "College undermatching, degree attainment, and minority students," *Paper presented at the 2018 AERA Annual Meeting.*

Karaagaoglu, Nilgun and Sevinc Yucecan, (2000). "Some behavioral changes observed among fasting subjects, their nutritional habits and energy expenditure in Ramadan," *International Journal of Food Sciences and Nutrition*, *51*(2), p. 125.

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

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

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

Kunda, Ziva, (1990). "The case for motivated reasoning.," *Psychological Bulletin*, *108*(3), p. 480.

Kuran, Timur, (2018). "Islam and economic performance: Historical and contemporary links," *Journal of Economic Literature*, *56*(4), pp. 1292–1359.

Lancet, (2009). "Ramadan: Health effects of fasting," *The Lancet*(9690).

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

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

Lee, Jaekyung, Lois Weis, Keqiao Liu, and Chungseo Kang, (2017). "Which type of high school maximizes students' college match? Unequal pathways to postsecondary destinations for students from varying high school settings," *The Journal of Higher Education*, *88*(4), pp. 529–560.

Leiper, John Beattie and AM Molla, (2003). "Effects on health of fluid restriction during fasting in ramadan," *European Journal of Clinical Nutrition*, *57*(S2), p. S30.

Li, Hongbin, Prashant Loyalka, Scott Rozelle, Binzhen Wu, and Jieyu Xie, (2015). "Unequal access to college in China: How far have poor, rural students been left behind?" *The China Quarterly*, *221*, pp. 185–207.

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

Little, Thomas S and Nannette B Dacus, (1999). "Looping: Moving up with the class.," *Educational Leadership*, *57*(1), pp. 42–45.

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

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

Loyalka, Prashant, Sean Sylvia, Chengfang Liu, James Chu, and Yaojiang Shi, (2019). "Pay by design: Teacher performance pay design and the distribution of student achievement," *Journal of Labor Economics*, *37*(3), pp. 621–662.

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

Majid, Muhammad Farhan, (2015). "The persistent effects of in utero nutrition shocks over the life cycle: Evidence from Ramadan fasting," *Journal of Development Economics*, *117*, pp. 48–57.

Manski, Charles F, (1993). "Identification of endogenous social effects: The reflection problem," *The Review of Economic Studies*, *60*(3), pp. 531–542.

Manski, Charles F and David A Wise, (1983). *College Choice in America*: Harvard University Press.

Mbiti, Isaac M, (2016). "The need for accountability in education in developing countries," *Journal of Economic Perspectives*, *30*(3), pp. 109–32.

Mbiti, Isaac, Karthik Muralidharan, Mauricio Romero, Youdi Schipper, Constantine Manda, and Rakesh Rajani, (2018). "Inputs, incentives, and complementarities in education: Experimental evidence from Tanzania," *NBER Working Paper No.24876*.

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

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

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

Mobius, Markus M, Muriel Niederle, Paul Niehaus, and Tanya S Rosenblat, (2011). "Managing self-confidence: Theory and experimental evidence,"Technical report, National Bureau of Economic Research.

Montgomery, James D, (1996). "Contemplations on the economic approach to religious behavior," *American Economic Review*, *86*(2), pp. 443–447.

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

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

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

Muralidharan, Karthik and Venkatesh Sundararaman, (2011). "Teacher opinions on performance pay: Evidence from India," *Economics of Education Review*, *30*(3), pp. 394–403.

Neal, Derek, (2011). "The design of performance pay in education," In *Handbook of the Economics of Education*, *4*: Elsevier, pp. 495–550.

Neal, Derek and Diane Whitmore Schanzenbach, (2010). "Left behind by design: Proficiency counts and test-based accountability," *The Review of Economics and Statistics*, *92*(2), pp. 263–283.

Netzer, Nick and Jean-Michel Benkert, (2018). "Informational requirements of nudging," *Journal of Political Economy*, *forthcoming*.

Nyhan, Brendan and Jason Reifler, (2010). "When corrections fail: The persistence of political misperceptions," *Political Behavior*, *32*(2), pp. 303–330.

Oosterbeek, Hessel and Bas van der Klaauw, (2013). "Ramadan, fasting and educational outcomes," *Economics of Education Review*, *34*, pp. 219–226.

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

Oreopoulos, Philip and Reuben Ford, (2019). "Keeping college options open: A field experiment to help all high school seniors through the college application process," *Journal of Policy Analysis and Management*, *38*(2), pp. 426–454.

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

Oster, Emily, Ira Shoulson, and E Dorsey, (2013). "Optimal expectations and limited medical testing: Evidence from Huntington disease," *American Economic Review*, *103*(2), pp. 804–30.

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

Page, Lindsay, Christopher Avery, and Jessica Howell, (2014). "A review of the role of college counseling, coaching, and mentoring on students' postsecondary outcomes," *College Board Research Brief*.

Page, Lindsay C and Hunter Gehlbach, (2017). "How an artificially intelligent virtual assistant helps students navigate the road to college," *AERA Open*, *3*(4), pp. 1–12.

Page, Lindsay C, Stacy S Kehoe, Benjamin L Castleman, and Gumilang Aryo Sahadewo, (2019). "More than dollars for scholars the impact of the dell scholars program on college access, persistence, and degree attainment," *Journal of Human Resources*, *54*(3), pp. 683–725.

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

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

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

Paulsen, Michael B, (2001). "The economics of human capital and investment in higher education," *The Finance of Higher Education: Theory, Research, Policy, and Practice*, pp. 55–94.

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

Perna, Laura W, Heather T Rowan-Kenyon, Scott Loring Thomas, Angela Bell, Robert Anderson, and Chunyan Li, (2008). "The role of college counseling in shaping college opportunity: Variations across high schools," *The Review of Higher Education*, *31*(2), pp. 131–159.

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

Peter, Frauke, C Katharina Spiess, and Vaishali Zambre, (2018). "Informing students about college: An efficient way to decrease the socio-economic gap in enrollment: Evidence from a randomized field experiment," *DIW Berlin Discussion Paper No. 1770.*

Plutzer, Eric and Michael Berkman, (2008). "Trends: Evolution, creationism, and the teaching of human origins in schools," *Public Opinion Quarterly*, *72*(3), pp. 540–553.

Prakhov, Ilya and Denis Sergienko, (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*.

Priest, Susanna Hornig, (2000). "US public opinion divided over biotechnology?" *Nature Biotechnology*, *18*(9), p. 939.

Redlawsk, David P, (2002). "Hot cognition or cool consideration? testing the effects of motivated reasoning on political decision making," *The Journal of Politics*, *64*(4), pp. 1021–1044.

Roderick, M, J Nagaoka, V Coca, E Moeller, K Roddie, J Gilliam, and D Patton, (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*.

Roderick, Melissa, Vanessa Coca, and Jenny Nagaoka, (2011). "Potholes on the road to college: High school effects in shaping urban students' participation in college application, four-year college enrollment, and college match," *Sociology of Education*, *84*(3), pp. 178–211.

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

Schofield, Heather, (2014). "The economic costs of low caloric intake: Evidence from India," *Unpublished Manuscript*.

Simon, Herbert A, (1955). "A behavioral model of rational choice," *The Quarterly Journal of Economics*, *69*(1), pp. 99–118.

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

Springer, Matthew G, John F Pane, Vi-Nhuan Le, Daniel F McCaffrey, Susan Freeman Burns, Laura S Hamilton, and Brian Stecher, (2012). "Team pay for performance: Experimental evidence from the round rock pilot project on team incentives," *Educational Evaluation and Policy Analysis*, *34*(4), pp. 367–390.

Stark, Rodney and Roger Finke, (2000). *Acts of Faith: Explaining the Human Side of Religion*: University of California Press.

Taber, Charles S and Milton Lodge, (2006). "Motivated skepticism in the evaluation of political beliefs," *American Journal of Political Science*, *50*(3), pp. 755–769.

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

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

Tuomi, Ilkka, (1999). "Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory," In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences*, pp. 1–12, IEEE.

Wang, Lei, Mengying Wu, Yaojiang Shi, Yazhuo Chen, Prashant Loyalka, James Chu, Kaleigh Kenny, and Scott Rozelle, (2017). "The effect of teacher looping on student achievement: Evidence from rural China," *REAP Working Paper 318*.

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*.

Wu, Xiaogang and Donald J Treiman, (2007). "Inequality and equality under chinese socialism: The hukou system and intergenerational occupational mobility," *American Journal of Sociology*, *113*(2), pp. 415–445.

Yenmez, M Bumin, (2018). "A college admissions clearinghouse," *Journal of Economic Theory*, *176*, pp. 859–885.

Ziaee, V, M Razaei, Z Ahmadinejad, H Shaikh, R Yousefi, L Yarmohammadi, F Bozorgi, and MJ Behjati, (2006). "The changes of metabolic profile and weight during ramadan fasting," *Singapore Medical Journal*, *47*(5), p. 409.