

Survey Experiment

Xiaoyang Ye

xiaoyang.ye@princeton.edu

“[S]urvey experiments are more experiments than survey.”

Mutz, D. C. (2011). Population-based survey experiments. Princeton University Press.

Overview

- ① Why do we need survey experiments?
- ② What can we do with survey experiments?
 - ▷ Solicit beliefs
 - ▷ Randomize interventions
 - ▷ Simulate policy alternatives
- ③ Example: Free teacher education
- ④ Practice example

Introduction

Potential outcomes

- *Factual vs. Counterfactual*

$$Y_i = T_i \cdot Y_i(1) + (1 - T_i) \cdot Y_i(0)$$

- ▷ T_i : a dummy variable indicating whether individual i receives treatment ($T_i = 1$) or not ($T_i = 0$)
- ▷ $Y_i(1)$: the outcome of individual i if she receives treatment
- ▷ $Y_i(0)$: the outcome of individual i if she does not receive treatment
- A valid causality question must involve well-defined causes (treatments, manipulations), and the counterfactuals should be unambiguously defined.

Fundamental problem of causal inference

- *Individual treatment effect*

$$\tau_i = Y_i(1) - Y_i(0)$$

- Causality is defined by potential outcomes, not by realized (observed) outcomes
- We can only observe one of the two potential outcomes
 - ▷ Missing data problem: Any statistical method dealing with treatment effects necessarily imputes the counterfactual part of the data.

Selection bias in observed outcomes

- Holland (1986):

$$\begin{aligned} & \mathbf{E}[Y_i(1)|T_i = 1] - \mathbf{E}[Y_i(0)|T_i = 0] \\ &= \underbrace{\mathbf{E}[Y_i(1)|T_i = 1] - \mathbf{E}[Y_i(0)|T_i = 1]}_{\tau_{ATT}} + \underbrace{\mathbf{E}[Y_i(0)|T_i = 1] - \mathbf{E}[Y_i(0)|T_i = 0]}_{\text{selection bias}} \end{aligned}$$

- Roy model:

Potential Outcomes: $Y_i(0) = \mathbf{X}_i\beta(0) + u_i(0)$

$$Y_i(1) = \mathbf{X}_i\beta(1) + u_i(1)$$

Selection/Assignment Mechanism: $\mathbf{1}_{\{T_i=1\}} = F(\mathbf{X}_i\gamma) + \epsilon_i$

▷ The identification is:

$$\mathbf{X}_i \perp (u_i(0), u_i(1), \epsilon_i)$$

① By knowledge of **Assignment Mechanism**

- ▷ Random assignment (RCT)
- ▷ Regression discontinuity (RD)

② By **Self-Selection**

- ▷ Difference-in-differences (DID)
 - Influence of “other factors” fixed
- ▷ Selection on unobservables and instrumental variables (IV)
 - Conditional on covariates, instrument “as good as randomly assigned” (uncorrelated with potential outcomes)
 - Another structural approach: Heckman selection model
- ▷ Selection on observables and matching (Matching)
 - Conditional on covariates, treatment “as good as randomly assigned”

RCT: The gold standard

- An experiment
 - ▷ The observation of units after, and possibly before, a randomly assigned intervention in a controlled setting, which tests one or more precise causal expectations.
- Key elements
 - ▷ Treatment
 - ▷ Control
 - ▷ Random assignment
 - independent of potential outcomes
 - independent of all confounding factors
- Don't just do an experiment
 - ▷ Theoretically motivated
 - ▷ Test important questions

Conducting an experiment in different places

- ① Lab experiment
 - ▷ Treatment in a controlled research environment
- ② Field experiment
 - ▷ Treatment in real life
- ③ Survey experiment
 - ▷ Treatment in a survey

Protocol

for how to design, implement, and analyze an experiment

1 Theory/hypotheses

- ▷ What is your research question?
- ▷ Connect the experiment to extant literature and the real world
- ▷ Testable hypotheses derived from theory
 - Expectations about differences in outcomes from the experiment

2 Design

- ▷ We test hypotheses by comparing experimental factors/conditions
 - presence/absence; levels/doses; qualitative variations
- ▷ Design considerations
 - Single factor vs. crossed designs
 - Internal validity (attrition, test effect, non-response, spillover)
 - Conclusion validity (power, content, construct, predictive)

Protocol

for how to design, implement, and analyze an experiment

3 Sampling

- ▷ Sample size
 - Statistical power and minimal detectable effect size (MDES)
 - Power: Probability of rejecting the null when a causal effect exists
- ▷ External validity
 - Setting, unit, treatment, outcome
 - Generalization across contexts and populations, replications

4 Implementation

- ▷ Covariates
 - Plan for use in advance, measure them well
- ▷ Balance in covariates
 - Stratified randomization (by covariates) as much as possible
- ▷ Moderation
 - Manipulate the moderator, or block on the moderator, or run interaction term regressions
- ▷ Mediation
 - Manipulate the mediator, or manipulate the mediator for some, or observe the mediator

Protocol

for how to design, implement, and analyze an experiment

5 Analysis

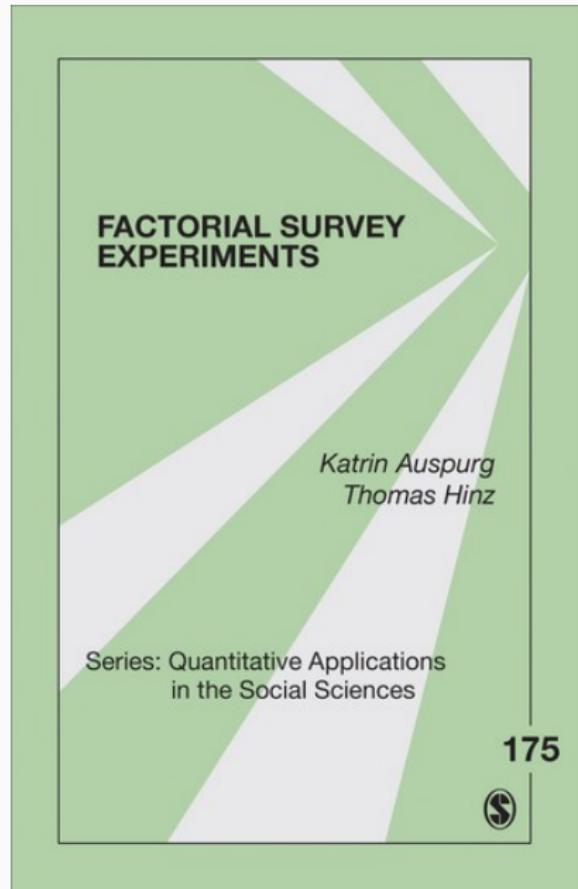
- ▷ Experimental inference
 - Parametric: t-test, ANOVA, regression
 - Non-parametric: build a randomization distribution (permutation)

6 Broken experiments

- ▷ Attrition
- ▷ Non-compliance (failure to treat or control gets treated)
- ▷ Missing data
- ▷ Some best practices
 - Pilot pretest, manipulation checks, placebo tests, non-equivalent outcomes

Survey experiment

Recommended reading



2. Problems and Techniques

Experiments in the Wording of Questions

THE IMPORTANCE of the wording of questions in public opinion polls was emphasized by Elmo Roper in the last issue of the QUARTERLY. In his article Mr. Roper stated that he had offered to test through his organization alternate wordings suggested to him by the QUARTERLY. The report below summarizes the first experiment on this problem.

Alternate wordings on two questions of current interest were chosen. In the first question one form used President Roosevelt's name, the other did not. In the second question Hitler's name was mentioned in one form but not in the other.

1(a). Do you approve of Sumner Welles' visit to European capitals?

1(b). Do you approve of President Roosevelt's sending Sumner Welles to visit European capitals?

2(a). Do you think the U.S. should do more than it is now doing to help England and France?

2(b). Do you think the U.S. should do more than it is now doing to help England and France in their fight against Hitler?

Two representative samples of the population, based on the criteria used by Roper for the *Fortune* surveys, were selected. Questions 1(a) and 2(b) were asked one group, questions 1(b) and 2(a) were asked the other group. Each group contained about 1550 persons. The survey was made during March 1940.

Welles' Visit

When President Roosevelt's name is used in connection with Mr. Welles' visit, two significant differences in total response are revealed: more people have opinions and, although the percentage of people who approve of the visit remains identical, more people disapprove of it when its initiation is attributed to Roosevelt. The percentages are:

	App.	Disapp.	No op.
1(a) [without Roosevelt]	43%	25%	32%

1(b) [with Roosevelt]	43	31	26
-----------------------	----	----	----

The President's name in this particular context, then, seemed to bring

Vignette

Example

- Theory
 - ▷ Economic information may close aspiration disparities for postsecondary education across socioeconomic, ethnic, and partisan divides
- Vignette/Manipulation
 - ▷ (1) Information about returns to a 2-year and 4-year degrees
 - ▷ (2) Information about net costs
 - ▷ (3) Information about both returns and costs
 - ▷ Control
- Outcome
 - ▷ “Would you want your child to go to a community college to earn a 2-year degree, a university to earn a 4-year degree, or neither?”

Cheng, A., & Peterson, P. E. (2019). Experimental Estimates of Impacts of Cost-Earnings Information on Adult Aspirations for Children's Postsecondary Education. *The Journal of Higher Education*, 90(3), 486-511.

Experimental Design and Survey Questionnaire

Table B1: Main Survey and Experimental Design

		No Costs Information	Providing Costs Information
No Returns Information	<p>Thinking about your oldest child under the age of 18 (If you had a child of college-going age), do you want that child to go to a community college to earn a two-year degree, a university to earn a four-year degree, or neither?</p>	<p>On average, it costs \$14,210 per year to complete a four-year degree at an in-state public university, while it costs \$7,620 per year to complete a two-year degree at a local community college. These are average costs (including tuition, fees, and room and board) after deducting the amount that students typically receive in scholarships and grants.</p> <p>Thinking about your oldest child under the age of 18 (If you had a child of college-going age), do you want that child to go to a community college to earn a two-year degree, a university to earn a four-year degree, or neither?</p>	
Providing Returns Information	<p>On average, students completing a four-year degree earn \$61,400 each year over the course of their working lives, while those completing a two-year degree earn \$46,000 each year over the course of their working lives.</p> <p>Thinking about your oldest child under the age of 18 (If you had a child of college-going age), do you want that child to go to a community college to earn a two-year degree, a university to earn a four-year degree, or neither?</p>	<p>On average, students completing a four-year degree earn \$61,400 each year over the course of their working lives, while those completing a two-year degree earn \$46,000 each year over the course of their working lives.</p> <p>On average, it costs \$14,210 per year to complete a four-year degree at an in-state public university, while it costs \$7,620 per year to complete a two-year degree at a local community college. These are average costs (including tuition, fees, and room and board) after deducting the amount that students typically receive in scholarships and grants.</p> <p>Thinking about your oldest child under the age of 18 (If you had a child of college-going age), do you want that child to go to a community college to earn a two-year degree, a university to earn a four-year degree, or neither?</p>	

Survey experiment designs

① Question wording designs

- ▷ e.g., "Free Teacher Education" vs. "Publicly-Funded Teacher Education"

② Question order designs

- ▷ Asking different pre-outcome questions (e.g., value salience) question-as-treatment

③ Vignettes

- ▷ A "vignette" is a short text describing a situation
- ▷ Vignettes are probably the most common survey experimental paradigm, after question wording designs
- ▷ Take many forms and increasingly encompass non-textual stimuli

④ Non-textual designs (e.g., image, audio, video)

⑤ Task designs (e.g., writing something)

Factorial design

An experiment can have any number of conditions

- Three types
 - ▷ Multiple conditions in a single factor
 - ▷ Multiple fully crossed factors
 - ▷ Partially crossed factors (“fractional factorial”)
- Consideration: Required sample size
 - ▷ 1 factor in 2 conditions (Treatment and Control) design: n
 - ▷ 4*4 design: n*16
- Consideration: Sensitive questions
 - ▷ List experiments
 - ▷ Randomized response
- Consideration: Conjoint analysis
 - ▷ Force-choice designs by comparing profiles of many features (revealed preferences)

Conducting a survey (experiment) in different places

- Face-to-face
- Computer Assisted Self-Interviewing
- Computer Assisted Telephone Interviewing
- Paper-and-pencil Self-Interviewing
- Web-based (or mobile) Self-Interviewing

Survey experiment protocol

- Sampling
 - ▷ What is your target population?
 - ▷ How do you sample and contact them?
 - ▷ Do you need a representative sample?
- Control
 - ▷ Surveys are less controlled (Setting, broader context, engagement) and more private than lab
 - ▷ Respondents decide when to complete study
- Questionnaire design
 - ▷ How do we measure constructs?
 - ▷ How do respondents understand those measures?
 - ▷ How do we hold respondents' attention?
 - ▷ Ethical considerations (esp. deception)

- Researchers have obligations to attempt to:
 - ▷ Minimize risk to participants
 - ▷ Maximize benefits to human knowledge
 - ▷ protect the privacy of personal data
 - ▷ Fairly and objectively report their research
- These rules vary to some extent across contexts

Recommended reading



Annual Review of Political Science

Some Advances in the Design of Survey Experiments

Paul M. Sniderman

Department of Political Science, Stanford University, Stanford, California 94305, USA;
email: paulms@stanford.edu

- Experimental design and hypothesis testing
 - ▷ Cross-category comparisons (vs. within-category)
 - ▷ Null by design
 - ▷ Explication
- Conjoint designs and the analysis of multidimensional choice
 - ▷ Causal inference in conjoint designs
 - ▷ Sequential factorials
 - ▷ Cultural pluralism
- Validation regimes
 - ▷ Parallel studies (replication)
 - ▷ Designing in pairs (reproducibility)
 - ▷ Splicing

Example: Future teachers

Recruiting the Very Best Future K-12 Teachers Using Free College

Xiaoyang Ye
Princeton University

2020 AEFP

The policy problem

- Recruiting and retaining high-quality teachers is essential to K-12 education
- Domestically and internationally, many school districts experience teacher shortages in underserved areas
- One policy option: **Financial aid incentives** tied to service commitment
 - ▷ e.g., federal TEACH grant and teacher loan forgiveness program
- Both policy designs and behavioral barriers may make these incentives ineffective

This paper

- The Free Teacher Education policy in China from 2007
 - ▷ In 6 most selective teachers colleges (#10,000 enrollment/year)
 - ▷ Full tuition waivers (“free college”) + stipends
 - ▷ Full-time teaching for 10 years after graduation

This paper

- The Free Teacher Education policy in China from 2007
 - ▷ In 6 most selective teachers colleges (#10,000 enrollment/year)
 - ▷ Full tuition waivers (“free college”) + stipends
 - ▷ Full-time teaching for 10 years after graduation
- ① Evaluation of the policy effects on college-major choice
 - Data from one state, 2001-2018
 - Event study design
- ② A survey experiment to test different policy designs and measure student preference
 - Service years: 0 vs. 6 vs. 10
 - Program name: “Free” vs. “Publicly-funded”

Background

The Free Teacher Education policy

- Teacher education in China
 - ▷ In early 20th century, closely following the U.S. model

The Free Teacher Education policy

- Teacher education in China
 - ▷ In early 20th century, closely following the U.S. model
 - ▷ Before late 1990s, free teacher education at secondary and post-secondary levels
 - ▷ 1999-2006, teachers colleges charged similar tuition and fees as other colleges of the same selectivity level

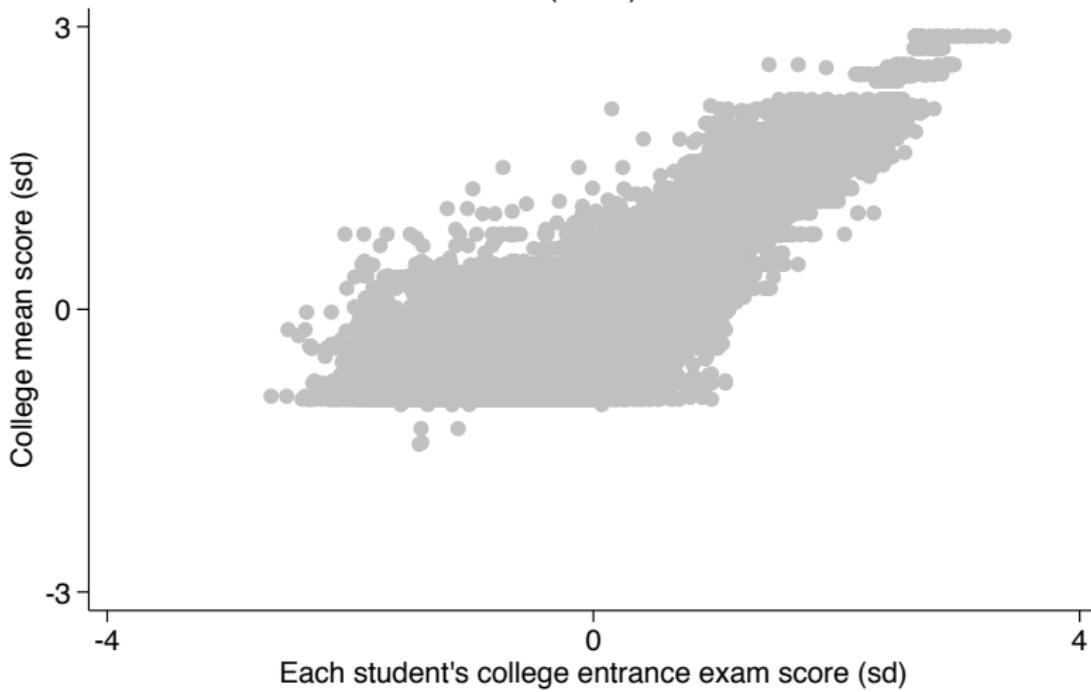
The Free Teacher Education policy

- Teacher education in China
 - ▷ In early 20th century, closely following the U.S. model
 - ▷ Before late 1990s, free teacher education at secondary and post-secondary levels
 - ▷ 1999-2006, teachers colleges charged similar tuition and fees as other colleges of the same selectivity level
 - ▷ 2007-2017, the national **Free Teacher Education** policy
 - ▷ 2018, modification to the FTE policy

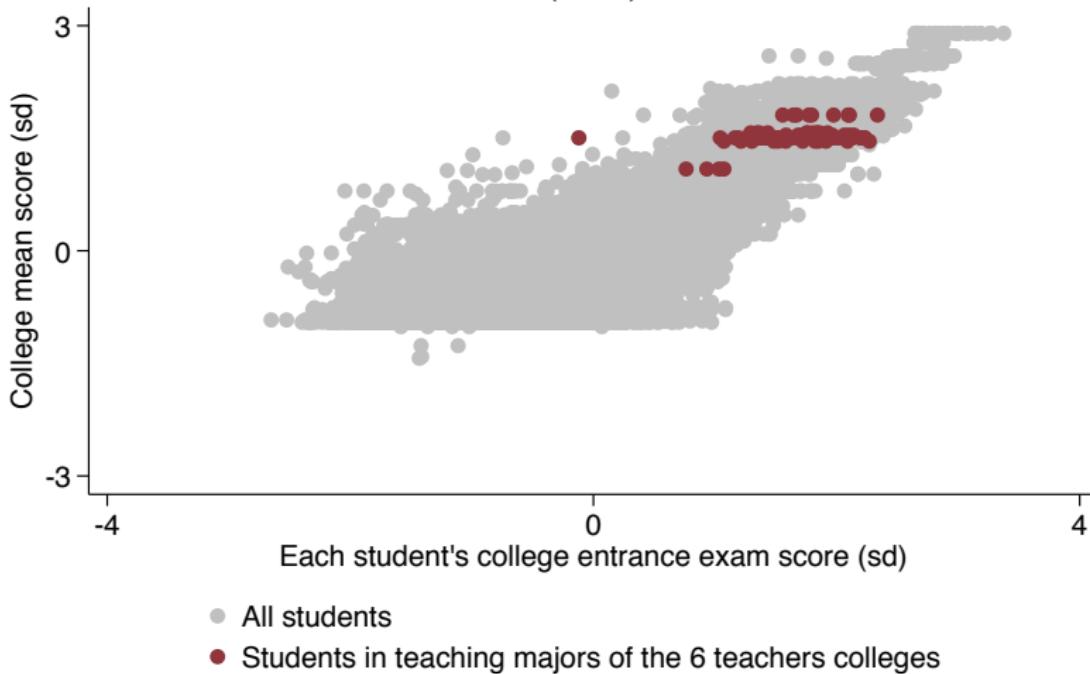
The Free Teacher Education policy

- The six top teachers colleges
 - ▷ Beijing Normal University
 - ▷ East China Normal University
 - ▷ Central China Normal University
 - ▷ Southwest China Normal University
 - ▷ Northeast China Normal University
 - ▷ Shaan'xi Normal University

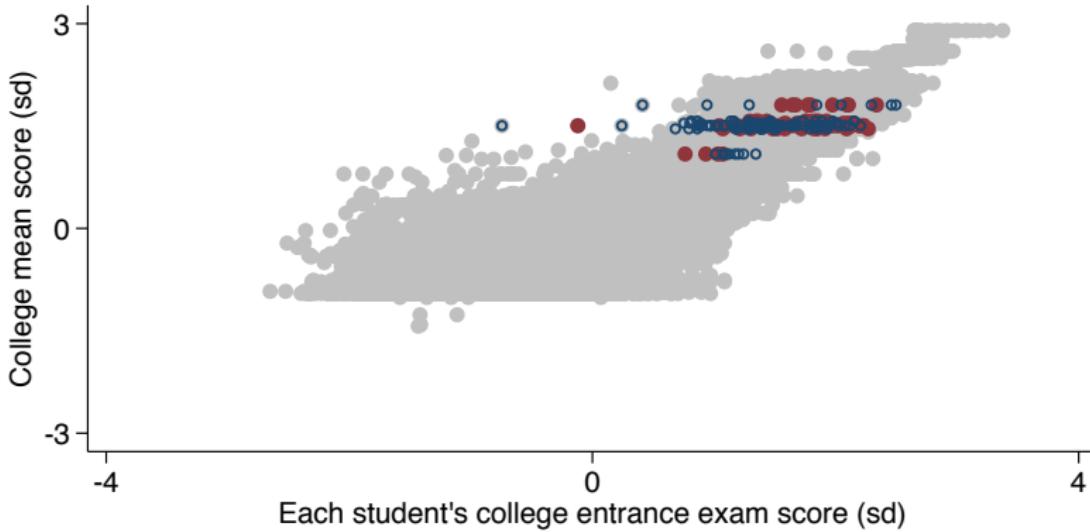
Distribution of college freshmen by mean entrance score (2006)



Distribution of college freshmen by mean entrance score (2006)

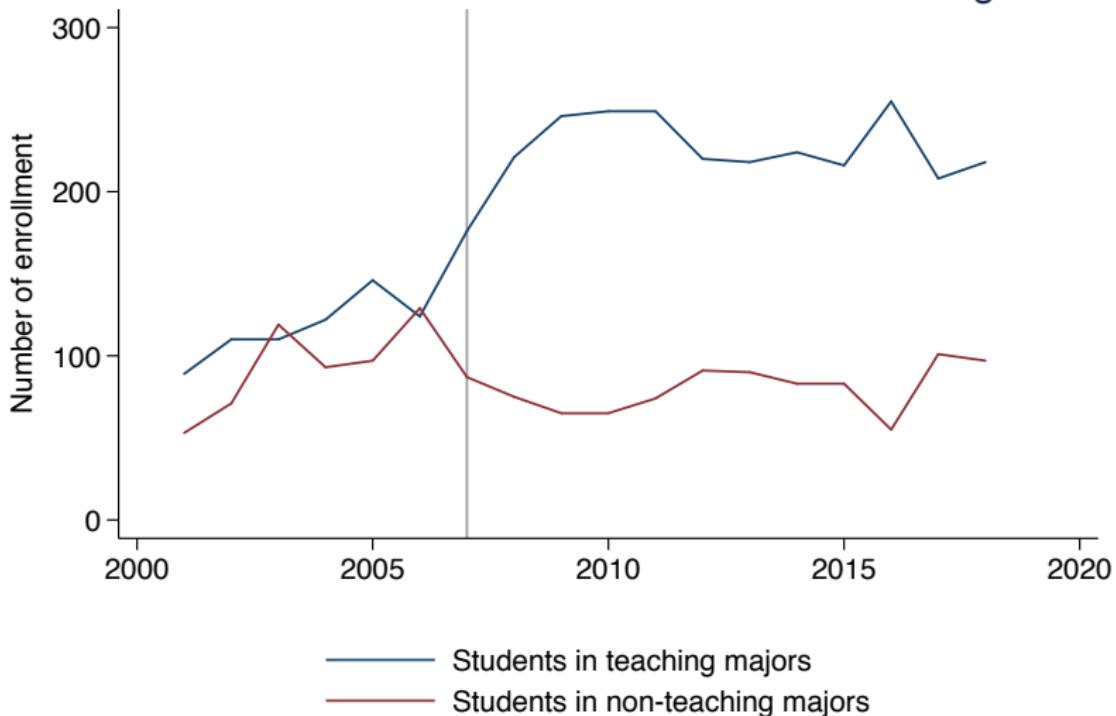


Distribution of college freshmen by mean entrance score (2006)



- All students
- Students in teaching majors of the 6 teachers colleges
- Students in non-teaching majors of the 6 teachers colleges

Freshman enrollment at the 6 teachers colleges



The Free Teacher Education policy from 2007

- To attract academically talented students into teaching
 - ▷ Free college + stipends
 - ▷ 10-year service commitment in the public school system
 - In their home states (low-income)
 - State centralized application and hiring
 - First two years in low-income schools
- To increase the attractiveness of the FTE policy (2018)
 - ▷ Service commitment
 - 10 years -> 6 years
 - 2 years in low-income schools -> 1 year
 - ▷ Name
 - “Publicly-funded teacher education”

In theory

- College choice - utility maximization
 - ▷ Financial aid incentives as increased benefits and expanded budget constraint
 - ▷ Teaching commitment as (positive or negative) long-term returns
- Heterogeneous beliefs and behavioral barriers
 - ▷ Gender gap: career preference, social norm
 - ▷ SES gap: information, guidance, inattentive
- “Brand name” effects
 - ▷ “Free” vs. “Publicly-funded”
 - ▷ “Scholarship” vs. “Aid.” (Avery & Hoxby, 2004)

Empirical strategy 1: Evaluating the FTE policy impacts

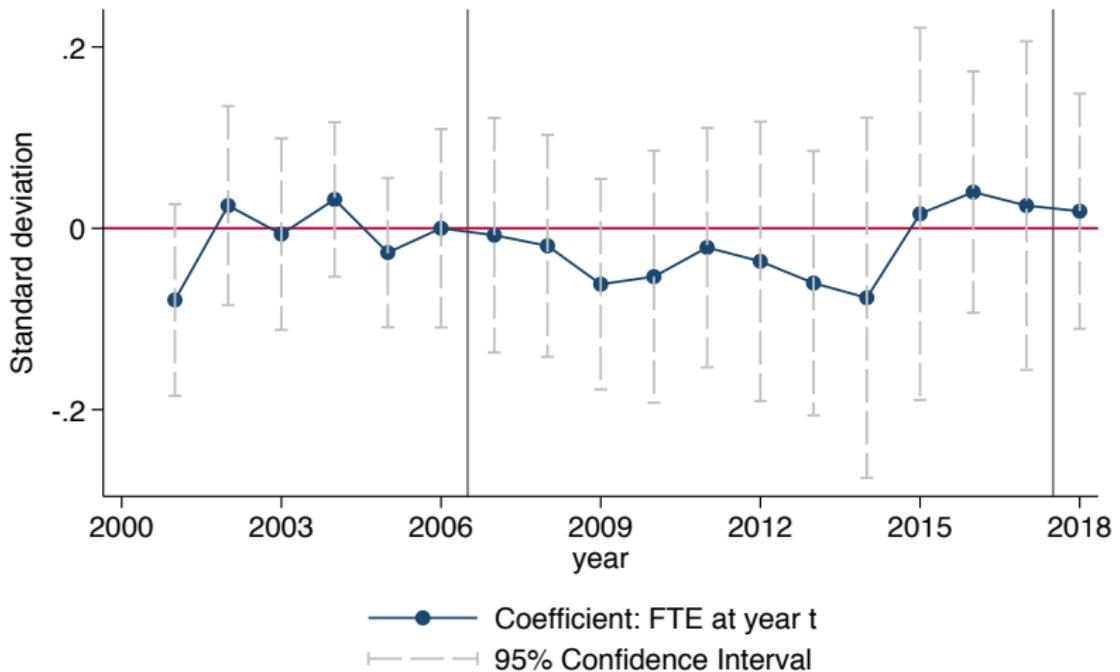
Generalized difference-in-differences

- Outcome of interest
 - ▷ How does the FTE policy change the distributions of enrolled students' achievements and demographics?

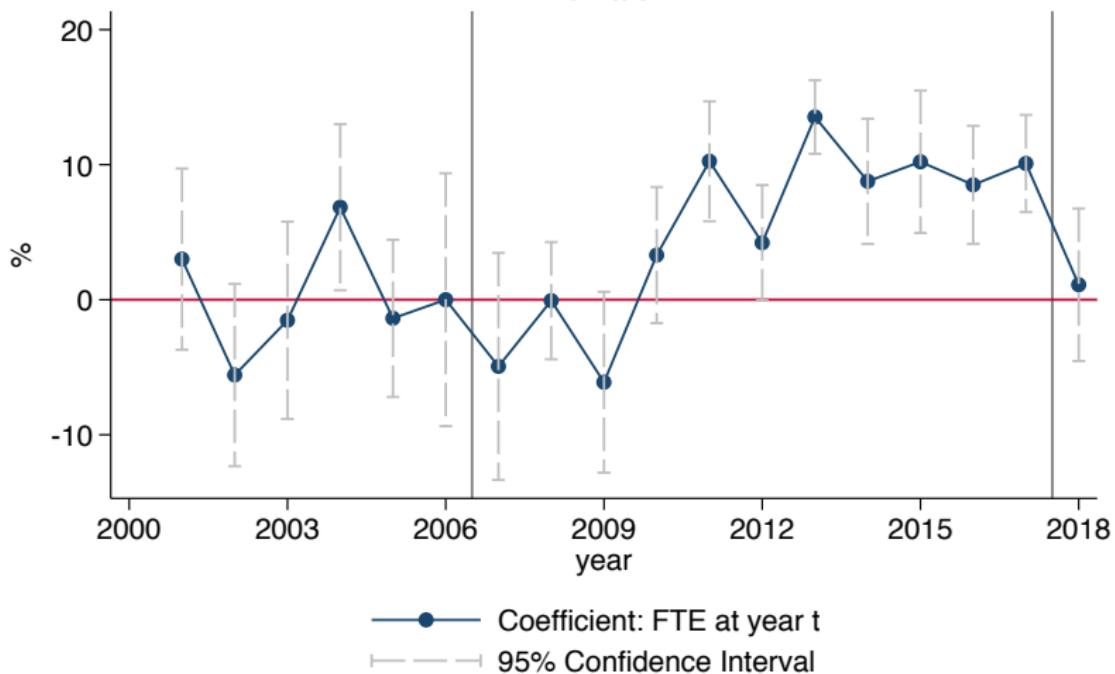
$$Y_{ijt} = \sum_{t \in \{2001-2018\}} \alpha_t \cdot FTE_{jt} \cdot Year_t + \beta \cdot X_{ijt} + \lambda_t + \theta_j + \epsilon_{ijt} \quad (1)$$

- FTE_{jt} equals 1 if college j provides free teacher education in year t

Impact of FTE on admitted student achievement

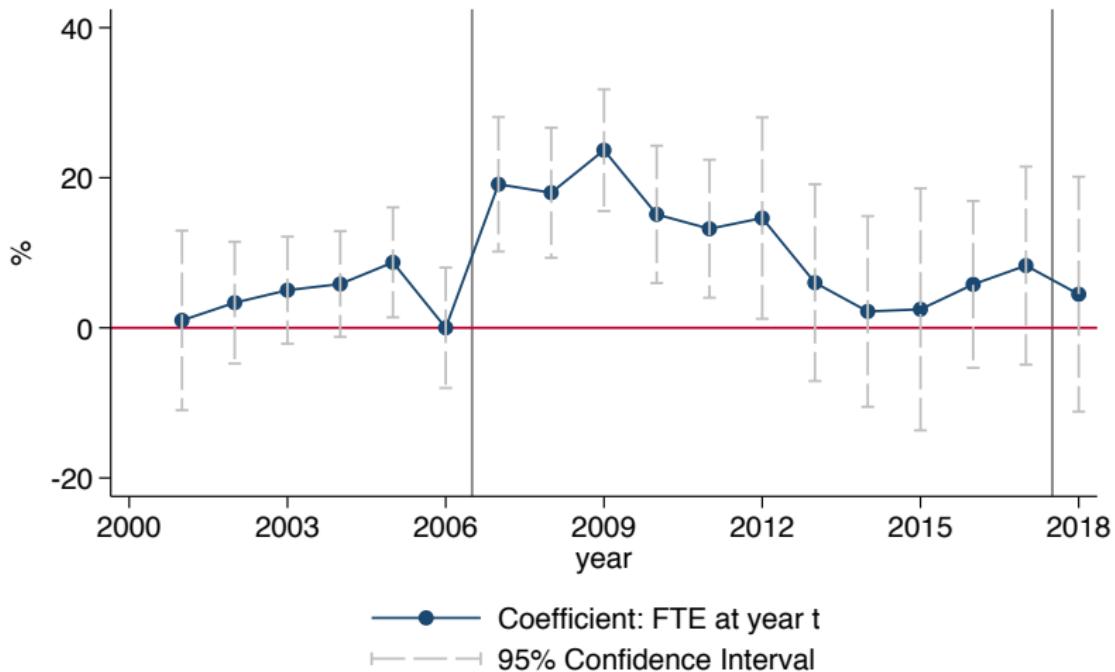


Impact of FTE on the percent of admitted students female



Impact of FTE on the percent of admitted students

rural



Empirical strategy 2: Measuring student preferences using a survey experiment

3*2 design

Treatment groups	Service Years	Name
1 (2007 Policy)	10	Free Teacher Education
2	6	Free Teacher Education
3	0	Free Teacher Education
4	10	Publicly-funded Teacher Education
5 (2018 Policy)	6	Publicly-funded Teacher Education
6	0	Publicly-funded Teacher Education

3*2 design

Treatment groups	Service Years	Name
1 (2007 Policy)	10	Free Teacher Education
2	6	Free Teacher Education
3	0	Free Teacher Education
4	10	Publicly-funded Teacher Education
5 (2018 Policy)	6	Publicly-funded Teacher Education
6	0	Publicly-funded Teacher Education

Survey experiment

- All senior students in a large urban high school of the state in Part 1
 - ▷ 1,190 students in 21 classes
 - ▷ Female 61%
 - ▷ From rural, poor families 30%

Survey experiment

- All senior students in a large urban high school of the state in Part 1
 - ▷ 1,190 students in 21 classes
 - ▷ Female 61%
 - ▷ From rural, poor families 30%
- A 30-minute survey in May 2018
 - ▷ Two weeks before the college entrance exam
 - ▷ Administered by classroom head teachers
 - ▷ Paperwork was randomized before distributing to students
 - ▷ No interactions between students during the survey

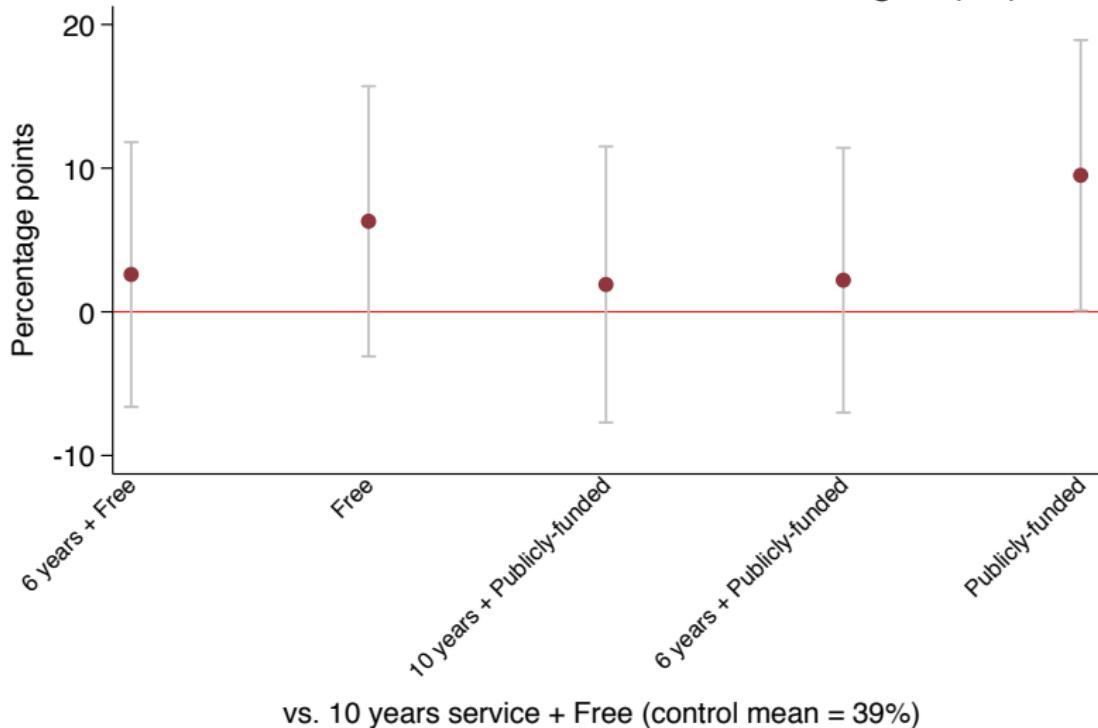
Within-class randomization

Treatment groups	Service Years	Name	Observations (%)
1 (2007 Policy)	10	Free	198 (16.6%)
2	6	Free	206 (17.3%)
3	0	Free	209 (17.6%)
4 (2018 Policy)	10	Publicly-funded	191 (16.1%)
5	6	Publicly-funded	197 (16.6%)
7	0	Publicly-funded	189 (15.9%)

$$Y_{ij} = \sum_{g \in \{2-6\}} \alpha_g \cdot Treatment_g + \beta \cdot X_{ij} + \theta_j + \epsilon_{ij}$$

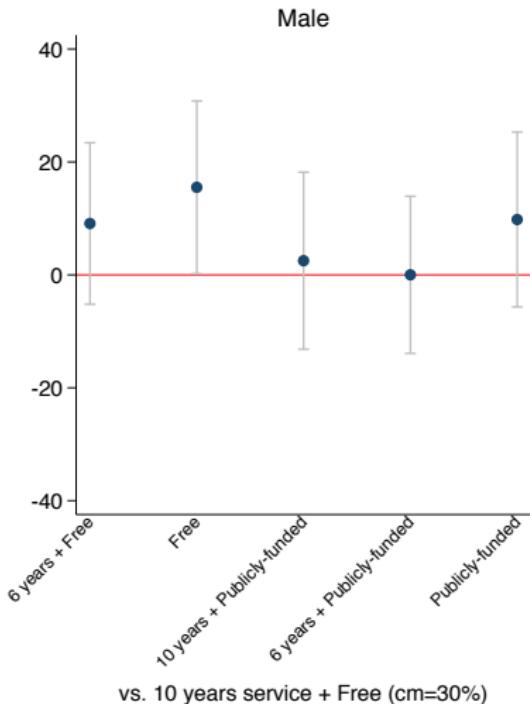
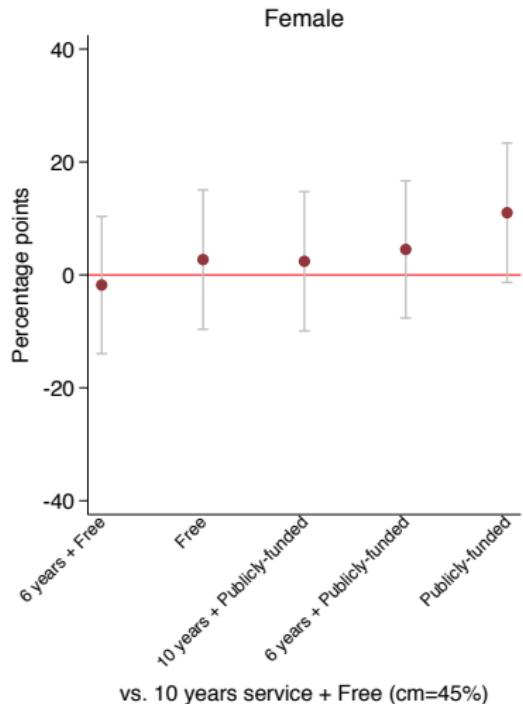
$$Y_{ij} = \sum_{g \in \{2-6\}} \alpha_g \cdot Treatment_g + \beta \cdot X_{ij} + \theta_j + \epsilon_{ij}$$

Outcome: Will consider teachers colleges (=1)



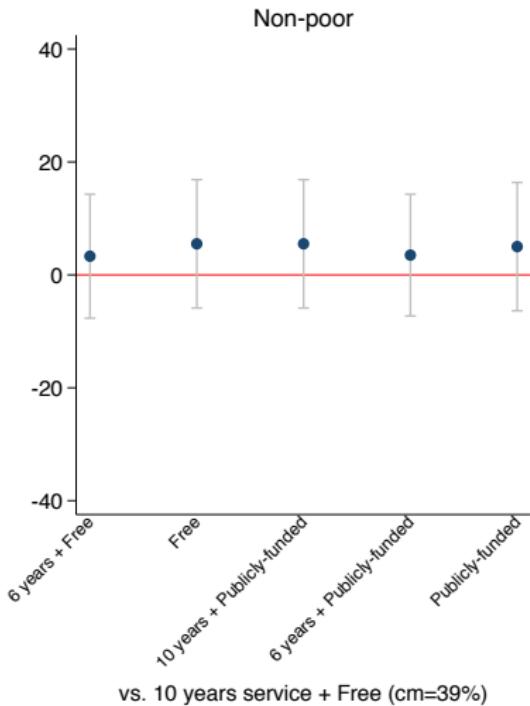
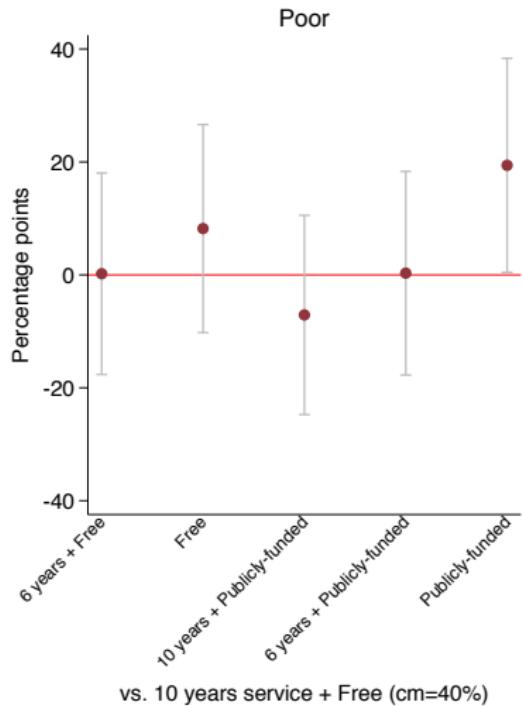
Gender and Poverty gaps

Outcome: Will choose a teachers college (=1)



Gender and Poverty gaps

Outcome: Will choose a teachers college (=1)



Summary

- The natural experimental evidence
 - ▷ Free college for teacher education may not sufficiently attract academically talented students given the attached service commitment
- The survey experimental evidence
 - ▷ Heterogeneous behavioral responses to the policy designs

Summary

- The natural experimental evidence
 - ▷ Free college for teacher education may not sufficiently attract academically talented students given the attached service commitment
- The survey experimental evidence
 - ▷ Heterogeneous behavioral responses to the policy designs
- Next steps
 - ▷ For this paper
 - Impacts on college-major application behaviors
 - Use national data of over 30 million students
 - ▷ For this project
 - Informational interventions in summer 2020
 - Impacts on college and career success

Practice example

Guidelines

- ➊ Cover everything in the protocol
 - ▷ Theory/hypotheses
 - ▷ Design
 - ▷ Sampling
 - ▷ Implementation
 - ▷ Analysis
 - ▷ Broken experiments
- ➋ Design and justify a budget
- ➌ Focus on anticipating challenges and your strategies for addressing them

如何利用大数据 优化贫困学生高等教育决策?



扫码填写问卷

“看滚滚长江东逝水
有好多事情要发生
——「酒之歌 The Spirit Song」”

- A. 对照组
- B. 为鼓励优秀高中毕业生报考师范专业，鼓励优秀青年长期从教，培养造就大批优秀教师和教育家，中国从2007年开始在北京师范大学、华东师范大学、东北师范大学、华中师范大学、陕西师范大学和西南大学等六所教育部属师范大学实施“公费师范生”项目，随后扩展到其他师范大学。
- C. 中国从2007年开始在北京师范大学、华东师范大学、东北师范大学、华中师范大学、陕西师范大学和西南大学等六所教育部属师范大学实施“免费师范生”项目，随后扩展到其他师范大学。学费、住宿费全免，在校期间发放生活费补贴。免费师范生承诺本科毕业后回生源所在省份从事中小学教育10年以上。
- D. 中国从2007年开始在北京师范大学、华东师范大学、东北师范大学、华中师范大学、陕西师范大学和西南大学等六所教育部属师范大学实施“公费师范生”项目，随后扩展到其他师范大学。学费、住宿费全免，在校期间发放生活费补贴。公费师范生承诺本科毕业后回生源所在省份从事中小学教育6年以上。

Online survey experiment

是否会考虑填报师范生专业？

控制组

免费师范生(免费、十年)

免费师范生

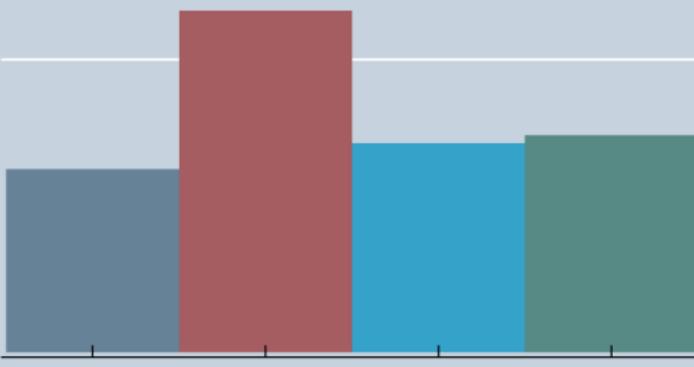
公费师范生(公费、六年)

20

比例 (%)

10

0



Summary

- Experiments are mostly about design, not analysis
- Careful but often simple design can generate potentially powerful and novel insights
- Learning outcomes
 - ① Fundamental problem of causal inference
 - ② Protocol of an experimental study
 - ③ Survey experiment for: beliefs, interventions, simulations
 - How to design survey experiments that speak to research questions and theories
 - Identify practical issues that arise in the implementation of survey experiments and evaluate how to anticipate and respond to them

Thanks!