Training Policymakers in Econometrics

Experimental Evidence from Two Field Experiments

Sultan Mehmood[1], Daniel Chen[2], Shaheen Naseer[3]

ASSA 2023 Session on Evidence Use in Policymaking

January 2023

[1]New Economic School of Moscow
[2]Toulouse School of Economics
[3]Lahore School of Economics

Motivation

- Over the last half century Economics has gone through a paradigm shift (Angrist and Pischke 2010)
  - The Credibility Revolution
  - 2019, 2021 Nobel Prizes
  - But it is an 'old' idea of "taking the con out of econometrics" (Leamer 1983)
- Broadly speaking, the credibility revolution focuses on:
  1. Causality
  2. Policy
- This makes it extremely relevant for policymakers

The Problem

- But the practitioners or policymakers may not respond to evidence.
- A large body of literature finds that policymakers are highly averse to shifting their beliefs and policy (Baekgaard et al. 2019; Banuri et al. 2019; Vivalt and Coville 2021; Bell and Toma 2021; Wang and Yang 2021; DellaVigna et al. 2022)
- Sticking to priors and being inattentive to evidence may stymie the implementation of good policies and hurt economic development (Kremer, Rao, Schilbach 2019)

Research Questions

- How can we make policymakers more receptive to evidence?
  - Will training them in concepts associated with the credibility revolution make them more likely to shift their beliefs?
  - Will it induce them to change their actual policy choices?
- **Can it increase State capacity?**
  - Will it increase use of causal versus correlational evidence?
  - Will it increase tax collection?

The Context

- Experiment 1 (**data collection and analysis completed**) Deputy Ministers of Pakistan
  - Policy experts who advice the PM, President and her Cabinet
  - "key wheels on which the entire engine of the state runs" (Government of Pakistan, 2019)
  - Chosen via a competitive examination:
    - About 200 are chosen from 15, 000 exam takers annually
    - Pass rate is about 1%
- Experiment 2 (**ongoing**, data collection completed, analysis pending)
  - Tax inspectors in Pakistan
    - They head the tax units (about 50 in a province)
    - Excise and Taxation Department (Government of Pakistan)
    - 230 tax officers (115 control and 115 metrics trained)

Experiment 1 - Overview

- Training deputy ministers in econometrics impacts
  1. **Attitudes** on importance of evidence
  2. **Willingness-to-Pay** for evidence
  3. **Policy** decisions

Experiment 1 - Magnitudes

- **Attitudes**:
  - 50% **increase** in importance of quantitative evidence
  - **No change** in importance of qualitative evidence
- **Stated Willingness-to-Pay**:
  - WTP for RCT Evidence **increases** by 300%
  - WTP for Correlational Evidence **decreases** by 50%
- **Actual Policy**:
  - Funding recommendations for policies with RCT evidence **doubles**
  - Amount of funding recommended **triples**
  - **No effect** on policies for which no RCT evidence was provided

Experiment 2 - Summary

- Randomly assign tax officers into metrics and control groups
  1. Pre-registered policy responses on tax reminder letters [**first-stage**]
  2. Pre-registered tax collection [**second-stage**]
  3. Can metrics training increase use of tax reminder policy and tax collection?
  4. **Additional**: Cross-randomized a handout (signal) to tax officers on evidence to adopt tax reminders policy
     - Tax reminders increases tax collection (RCT study)
     - Tax reminders increases tax collection (correlational study)

More Details 38

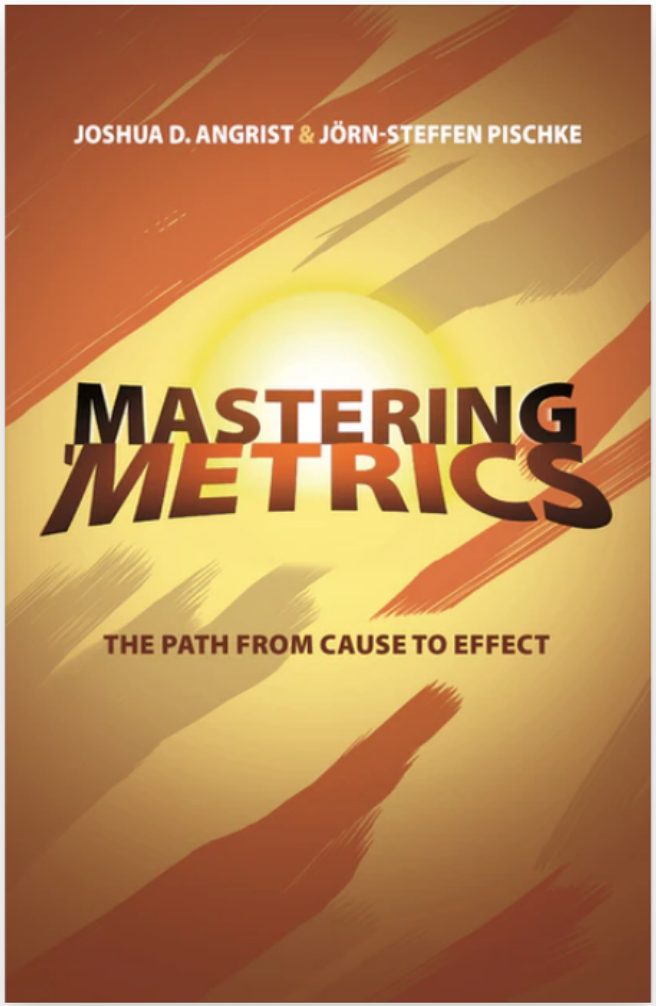
Related Literature and Contributions

- **Use of evidence in policy-making**
  - Policymakers may not use evidence due to many behavioral factors (Baekgaard et al. 2019; Banuri et al. 2019; Vivalt and Coville 2021; Bell and Toma 2021; Wang and Yang, 2021; DellaVigna et al. 2022)
    - ➤ Training policymakers in econometrics may alleviate some of the bottlenecks
- **Paradigm Shifts**
  - Pivots the literature on how and why paradigm shifts occur in science to study its consequences (Kuhn 1962, Shapin 1982, Merton 1973; Foucault 1970)
    - ➤ Study the causal effects of causal thinking
- **Economics and Financial Literacy**
  - Adds to the literature on economics and financial literacy (Cole et al., 2011, Lusardi et al., 2017, Sutter et al. 2020)
    - ➤ Study the impact of econometrics literacy training

Roadmap

1. Background
2. Data
3. Empirical Specification
4. Results
5. Conclusion
6. Appendix: Robustness Checks (balance, attrition, MHT, RI)

Metrics Training



CONTENTS

Placebo Training

Table of Contents

Intervention Details

- Ministers are randomly assigned to a metrics or a placebo workshop
  - Treated group: Mastering Metrics by Josh Angrist and Steve Pischke.
  - Placebo self-help book training: Mindsight by Daniel Siegel
- The ministers:
  - Write 1500 word essays summarising each chapter of the book
  - Another 1500 word essay on how they would apply the book's concepts in their policymaking
  - This is based on on recent scholarship on socio-emotional learning (Yeager et al., 2019, Nature)
  - These essays or assignments are rated in a competitive manner with cash awards and commemorative shields for top performers
  - Presentation of books' lessons within the treatment arm
  - Structured discussion within the treatment arm

Commemorative Shields and Vouchers

Panel A: Commemorative Shield

Note: The figure shows one of the commemorative shields presented to the deputy ministers.

Panel B: Gift Vouchers

Note: The figure shows cash gift vouchers at a luxury departmental store. The monetary amount is designated in Pakistan Rupees. The vouchers for the first three positions within each treatment arm and are worth about USD 150, USD 100 and USD 80, respectively.

Tax Officers Experiment

Randomized RCT Handout

BEHAVIORAL INTERVENTIONS IN TAX COMPLIANCE

Evidence from Guatemala

Policy Insights

- Sending Tax Notices or Tax Reminders increase tax collection.
- The individuals assigned Tax Reminder letters (notifications) increased tax compliance relative to those that were assigned no letters in control
- The effects are large and statistically significant.

Study Summary

In 2014, taxpayers (individuals and firms) who had failed to pay their income tax for the 2013 tax year Guatemala were randomly allocated to receive either no letter or tax reminder letters.

The randomized evaluation finds that although all letters increased the rate of declaration, the best performing treatments were a deliberate choice letters framing non-declaration as an intentional and deliberate choice.

These interventions increased the rate of payment as well as the average amount paid conditional on paying.

Tax Officers Experiment

Non-RCT Handout

EFFECT OF NOTICE

of Tax Warning, Notice of Tax Collection, and Tax Education Programs on Tax Compliance in West Sumatera and Jambi

Policy Insights

- Sending Tax Notices or Tax Reminders increase tax collection.
- Those who got Tax Reminder letters (notifications) increased tax compliance to those that did not get reminder letters.
- The effects are large and statistically significant.

Study Summary

This study conduct a panel regression analysis utilizing the administrative data from documentation study, and primary data through in-depth interviews with key informants and Supervision effort data consists of the number of Notice of Tax Warning and Notice of Tax Collection at West Sumatera and Jambi Regional Office.

Based on the findings and discussion, this study shows :

- This study effectively shows Notice of Tax Warning and Notice of Tax Collection effectively deliver law enforcement's message as a coercive power through supervision efforts.
- Tax education programs are practical to encourage

The Data

1. Experiment 1: Deputy Ministers
   - The Training Academy Course Records
   - The Federal Public Service Commission (FPSC)
   - Own Surveys and Academy's Policy Simulations
   - Performance Evaluation Reports
   - Sample of 190 Deputy Ministers out of 213 participated
2. Experiment 2: Tax Officers
   - External Audit Reports (ACRs, PRRs)
   - Administrative data from tax inspectors and Inland revenue services
   - 230 tax officers heading their respective units participated

OLS Estimation

$$Y_i = \theta + \alpha \text{ Metrics Assigned}_i + \mathbf{W'}_i\psi + \epsilon_i \quad (1)$$

- Subscript i is an individual deputy minister
- Y represents the respective outcome
- **W** is a vector of individual characteristics including choice of metrics book
- Clustering is done at the individual level
- α gives the causal impact of metrics training

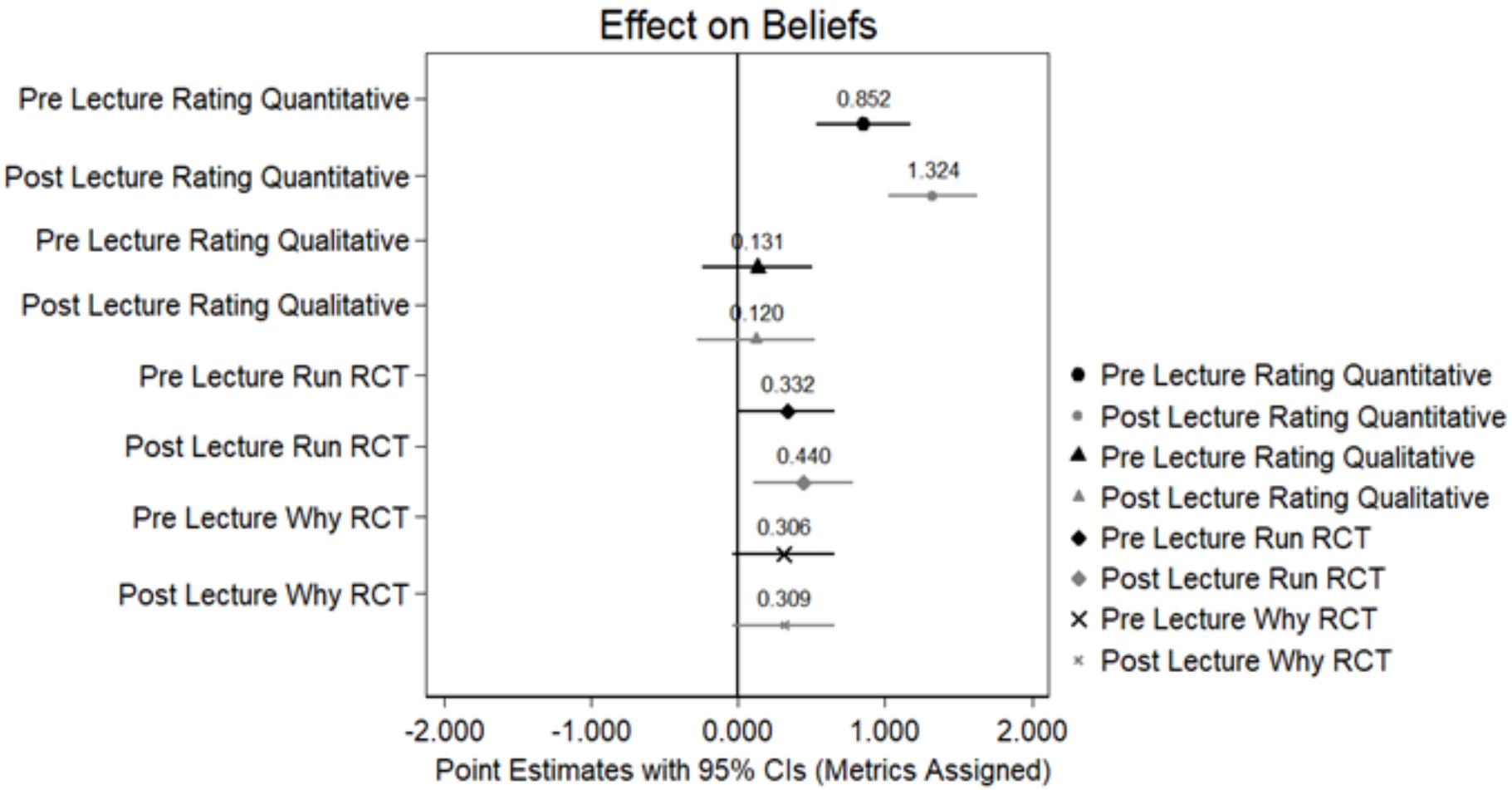
Partial and Full Metrics Training - Legend for Figures

- Partial Training = Pre-Lecture
  - Write a short summary of each chapter of the book
  - Write how the book's concepts could apply to your policymaking
- Full Training = Post Lecture
  - Video Lectures by authors: Josha Angrist and Daniel Siegelman
  - Structured discussion: "what are the key lessons?" "how will you apply these concepts in your career"
  - Presentation of the main lessons of the book
- Assignments and presentations were rated in a competitive manner with cash awards and commemorative shields awarded to top performers

Results I - Policy Attitudes - Original Units

	Pre-Lecture Rating Quantitative	Post-Lecture Rating Quantitative	Pre-Lecture Rating Qualitative	Post Lecture Rating Qualitative	Pre-Lecture Run RCT	Post-Lecture Run RCT	Pre-Lecture Why Run RCT	Post-Lecture Why Run RCT
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Metrics Assigned	0.912***	1.538***	0.136	0.122	0.167**	0 .220**	0.151*	0.153*
	(0.176)	(0.178)	(0.196)	(0.206)	(0.082)	(0.085)	(0.087)	(0.087)
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	190	190	190	190	190	190	190	190
Mean of dep. var. (placebo)	2.745	2.979	2.490	2.596	0.362	0.404	0.396	0.396

Results I - Policy Attitudes - Standardized Units



Stated Willingness-to-Pay

	Private Spending			*Public Spending*		
	Amount Randomized Trial	Amount Correlational Data	Amount Expert Bureaucrat	Amount Randomized Trial	Amount Correlational Data	Amount Expert Bureaucrat
	(1)	(2)	(3)	(4)	(5)	(6)
Metrics Assigned	2,063***	-1,020***	-1,986	1391308**	-35,274***	-2,048
	(587.5)	(340.4)	(1,390)	(665,160)	(13,033)	(34,090)
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	180	180	180	180	180	180
R-squared	0.217	0.202	0.104	0.094	0.217	0.202
Mean of dep. var. (placebo)	1539.453	2214.07	4490.008	928546.1	94136.72	115937.5

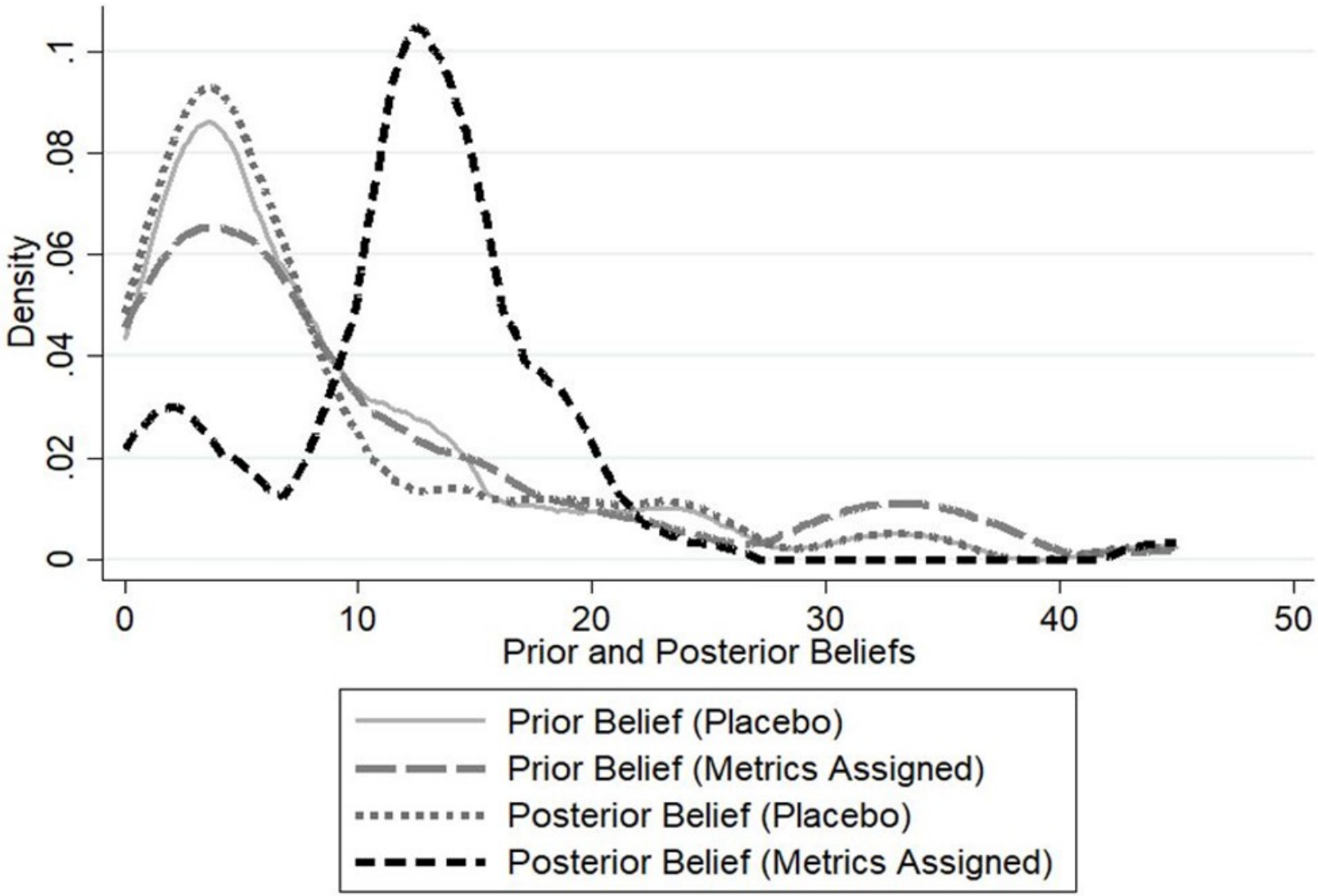
Initial and Posterior Beliefs on the Impact of Deworming

- We then collect prior and post-signal beliefs on the impact of two prominent policies that was to be rolled out by Government of Pakistan
  - Deworming in Schools
  - Computer Labs in Schools
- We then reveal a "signal" on the causal impact of Deworming in Schools on Long Run Incomes

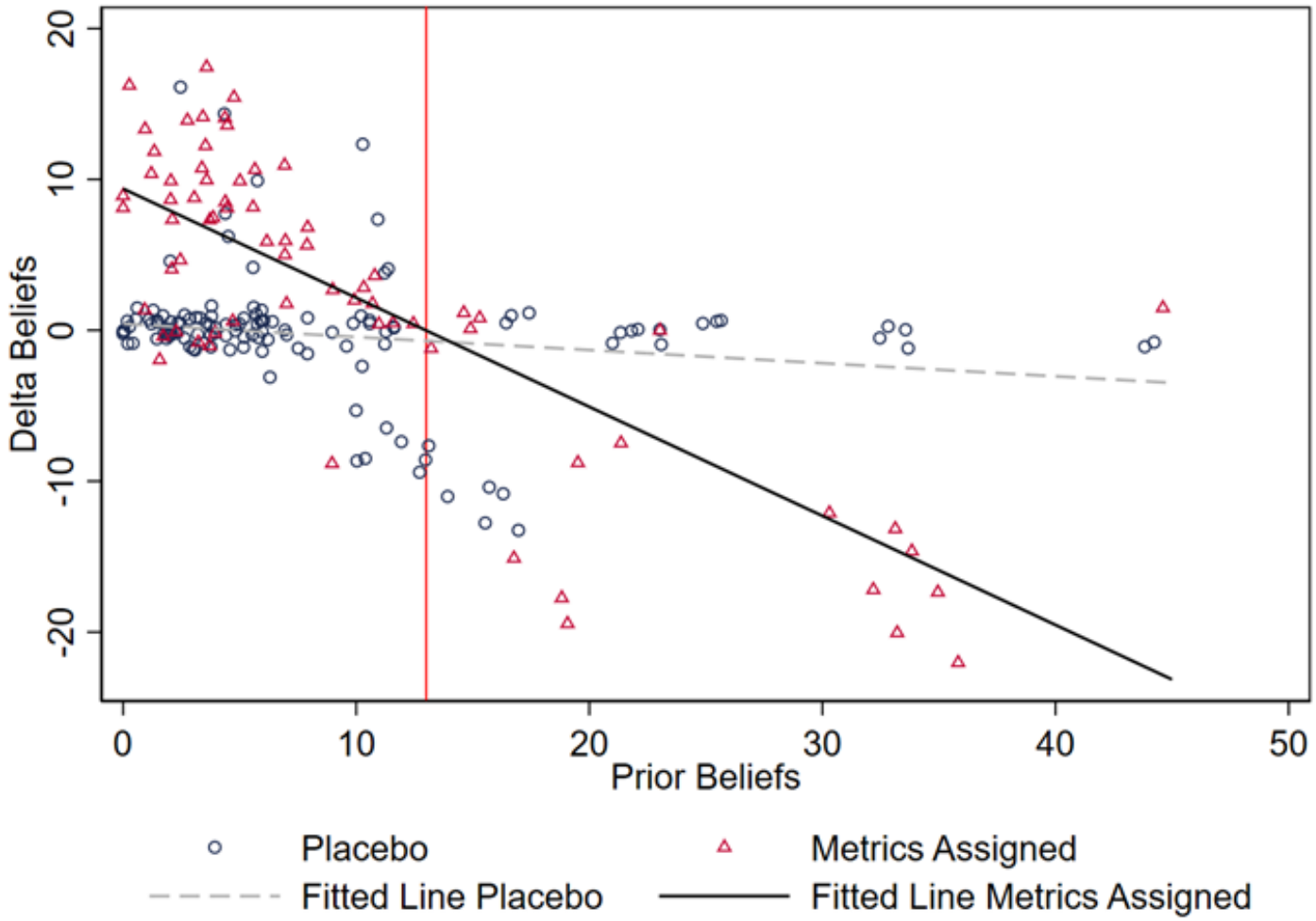
The Signal

Recent randomized evaluation finds deworming impacts on economic outcomes up to 20 years later. Individuals who received deworming experience up to 3 additional years of schooling, 14% increases in consumption expenditure, 13% increases in hourly earnings, 9% in non-agricultural work hours (Source: PNAS, 2021).

Initial and Post-Signal Beliefs on Deworming's Impact



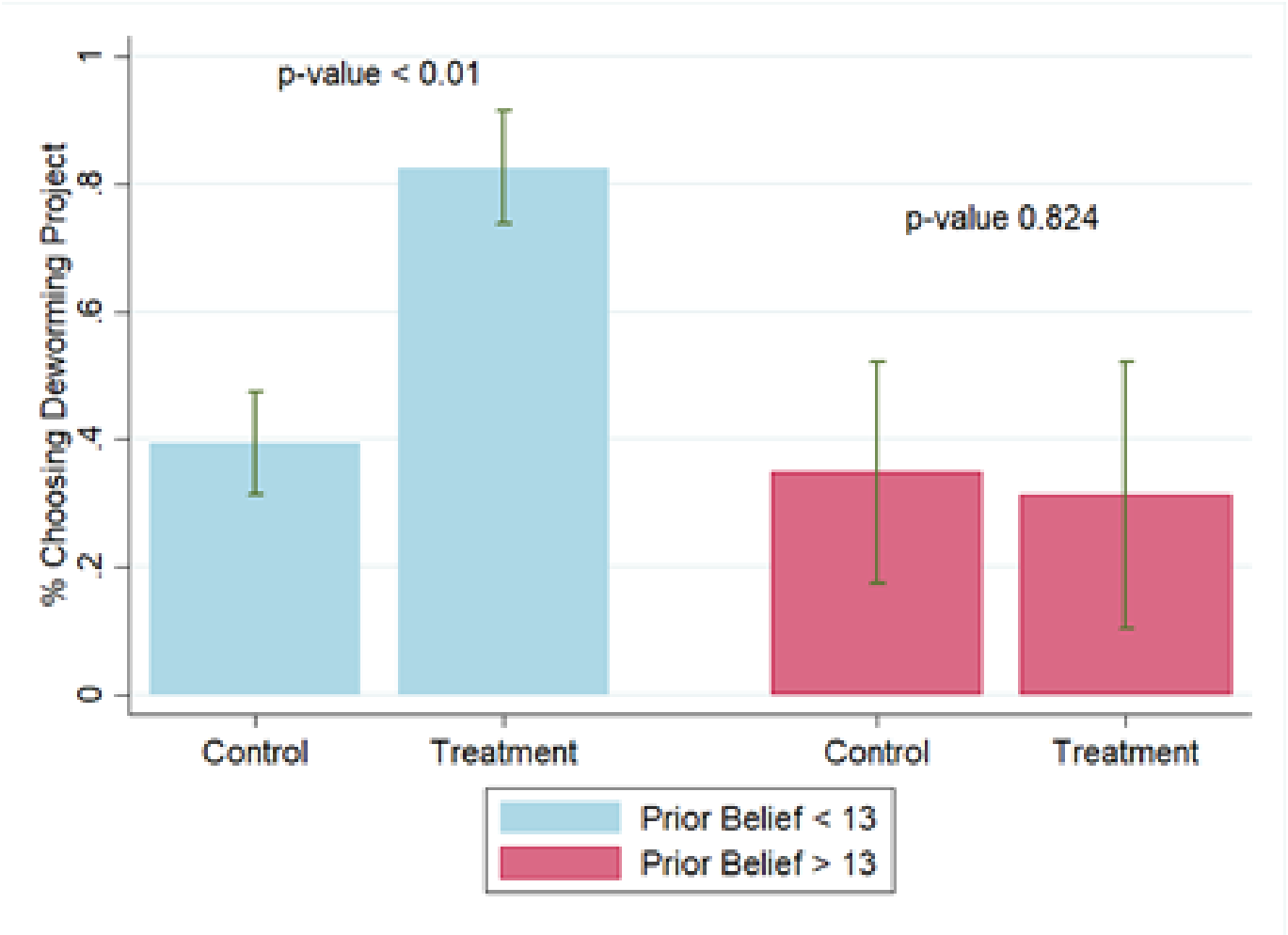
Initial and Shifts in Beliefs



Policy Choice facing Deputy Ministers

- Policy Choice
  - Deworming Policy in Schools
  - Computer Labs Policy in Schools

Impact on Stated Policy Choice



Impact on Policy

What about deputy ministers' policy responses in their official duties?

Impact on Actual Policy - Extensive and Intensive Margin

Budgetary Recommendations in Ministers' External Audit Reports

	Deworming Policy		***Orphanage Renovation Policy***		***School Renovation Policy***	
	Letter Sent	*Funds Recommended*	*Letter Sent*	*Funds Recommended*	*Letter Sent*	*Funds Recommended*
	(1)	(2)	(3)	(4)	(5)	(6)
Metrics Assigned	0.290***	401,888***	0.011	18,254	-0.053	-10,042
	(0.083)	(109,081)	(0.062)	(22,179)	(0.078)	(15,197)
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	190	190	190	190	190	190
R-squared	0.164	0.206	0.120	0.103	0.089	0.100
Mean of dep. var. (placebo)	0.174	171812.1	0.174	51073.83	0.262	41744.97

The Bottomline

When faced with policy choices having real reputational costs, implementation challenges and public budgetary constraints, treated policymakers choose policy for which there is causal evidence.

Experiment 2 with Tax Officers

- Does this replicate to tax officers?
  - Can econometrics training impact tax policy?
  - Can it impact tax collection?
  - How much metrics training allow people to be more attentive to different kinds of evidence?

Conclusion

- We find that training in the paradigm of the credibility revolution impacts:
  - Attitudes on importance of quantitative evidence
  - Willingness-to-Pay for experimental evidence
  - Policy decisions
- One year after the training, in their official duties, treated policymakers are:
  - Twice as likely to actually choose policies for which there is causal evidence
  - Triple the funding recommendations for such policies
- Impact of metrics training on:
  - State Capacity
  - Ability to distinguish between different types of evidence remains to be seen (**tax inspectors experiment**)

Thank You for Your Attention

- Twitter: **@mrsultan713**
- Email for more detailed questions/feedback: smehmood@nes.ru

Heterogeneity

1. Pre-treatment quantitative Test Scores
2. Treatment effect by Defiers and Never-Takers

Pre-treatment quantitative test scores

	Rating Quantitative		*Run RCT*		*Amount Randomized Trial (Private Spending)*		*Amount Randomized Trial (Public Spending)*	
Quantitative Scores	*Below Median*	*Above Median*	*Below Median*	*Above Median*	*Below Median*	*Above Median*	*Below Median*	*Above Median*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Metrics Assigned	1.735***	1.348***	0.166	0.168	2262.053**	2701.089**	1668811**	2883912**
	(0.220)	(0.422)	(0.114)	(0.153)	(1028.618)	(1114.788)	(637829)	(1186649)
Metrics Assigned Equal (p-value):	(1) = (2):	[0.501]	(3) = (4):	0.964	(5) = (6):	[0.555]	(7) = (8):	[0.482]
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	84	84	84	84	84	84	84	84
R-squared	0.592	0.448	0.212	0.510	0.402	0.295	0.167	0.285

Treatment effect by Defiers and Never-Takers

via Ministers' Click Behavior

	Private Spending		*Public Spending*	
	Amount Randomized Trial	Amount Randomized Trial	Amount Randomized Trial	Amount Randomized Trial
	(1)	(2)	(3)	(4)
Defiers X Metrics Assigned	-888.410		-2531651*	
	(1301.347)		(1367816)	
Never Takers X Metrics Assigned		-1987.019		-1304302
		(1837.863)		(1558128)
Metrics Assigned	2179.581***	2165.335***	1775415***	15028283**
	(648.553)	(641.327)	(635063.5)	(702559.8)
Defiers X Metrics Chosen	-428.640		3326010	
	(1207.932)		(2437943)	
Never Takers X Metrics Chosen		-643.626		2190478*
		(1385.195)		(1303170)
Defiers	-463.802		-3101800	
	(857.643)		(2201009)	
Never Takers		1756.372		-1250269
		(1887.535)		(1201626)

Econometrics Training and State Capacity

- A large body of evidence indicates that tax collection is key to build state capacity (Besley and Persson, 2009; Acemoglu et al., 2014; Gennaioli and Voth, 2015; Pomeranz and Vila-Belda, 2019; Bandiera et al., 2020)
  - Tax capacity is key ingredient to what makes an "effective state" (Bandiera et al., 2020; Olken et al., 2021)
- Theory of change:
  - Econometrics training would increase up-take of policy for which there is causal evidence (reminder letters)
  - Metrics trained tax officers may respond more to a signal of causal versus correlational evidence (first-stage)
  - Greater use of causal evidence will increase tax collection (second-stage)
- Back 8

Balance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Birth in political capitals	Income	Age	Education	Visited Foreign Country	PAS	PSP	Other groups	Pre-Treatment Written Assessment	Pre-Treatment Interview Assessment	Pre-Treatment Mathematics Assessment
Metrics Assigned	0.0528	-7,327	0.212	0.104	-0.00229	-0.0130	-0.0549	0.0235	0.960	2.208	0.0627
	(0.0902)	(4,601)	(0.395)	(0.0873)	(0.0712)	(0.0438)	(0.0348)	(0.0570)	(5.049)	(3.091)	(0.218)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	190	190	190	190	190	190	190	190	190	190	190
R-squared	0.047	0.165	0.253	0.225	0.070	0.630	0.453	0.645	0.484	0.227	0.017
Mean of dependent variable	0.324	34258.26	26.775	0.516	0.225	0.169	0.099	0.610	655.585	131.085	7.221

Attrition

	Attrition in Sample 1		*Attrition in Sample 2*	
	(1)	(2)	(3)	(4)
Metrics Assigned	-0.058	-0.053	0.001	0.008
	(0.045)	(0.047)	(0.031)	(0.035)
Individual Controls	No	Yes	No	Yes
Observations	213	213	213	213
R-squared	0.01	0.09	0.014	0.052
Mean of dep. var. (placebo)	0.091	0.091	0.06	0.06

Randomization Inference - I

	Pre Lecture Rating Quantitative	*Post Lecture Rating Quantitative*	*Pre Lecture Rating Qualitative*	*Post Lecture Rating Qualitative*	*Pre Lecture Run RCT*	*Post Lecture Run RCT*	*Pre Lecture Why Run RCT*	*Post Lecture Why Run RCT*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Metrics Assigned	0.912	1.538	0.136	0.122	0.166	0 .220	0.151	0.153
	(0.000) ***	(0.000) ***	(0.487)	(0.554)	(0.046) **	(0.011) **	(0.085) *	(0.080) *
	{0.000} ***	{0.000} ***	{0.491}	{0.533}	{0.075} *	{0.012} **	{0.104}	{0.096} *
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	190	190	190	190	190	190	190	190
Mean of dep. var. (placebo)	2.657	2.657	2.657	2.714	0.400	0.400	0.386	0.379

Randomization Inference - II

	Deworming Policy		*Orphanage Renovation Policy*		*School Renovation Policy*	
	Letter Sent	*Funds Requested*	*Letter Sent*	*Funds Requested*	*Letter Sent*	*Funds Requested*
	(1)	(2)	(3)	(4)	(5)	(6)
Metrics Assigned	0.656	0.714	0.0288	0.131	-0.122	-0.0808
	(0.001) ***	(0.001) ***	(0.866)	(0.412)	(0.498)	(0.510)
	{0.001} ***	{0.001} ***	{0.868}	{0.508}	{0.507}	{0.635}
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	190	190	190	190	190	190
R-squared	0.164	0.206	0.120	0.103	0.089	0.100

Multiple Hypothesis Testing

	Lecture Rating Quantitative	Deworming Letter	Deworming Funds	Amount Randomization Trial	Amount Correlation Data	Assessment Public Policy	Assessment Research Methods
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Metrics assigned	1.541	0.290	401,888	1391308	-35,274	0.505	0.799
p-value	(0.001) ***	(0.001) ***	(0.001) ***	(0.038) **	(0.007) ***	(0.003) ***	(0.001) ***
Sharpened q-value	[0.002] ***	[0.002] ***	[0.002] ***	[0.009] ***	[0.004] ***	[0.002] ***	[0.002] ***
FWER p-value	{0.002} ***	{0.002} ***	{0.002} ***	{0.175}	{0.002} ***	{0.179}	{0.002} ***
Observations	190	190	190	180	180	190	190
R-squared	0.425	0.164	0.206	0.094	0.148	0.071	0.218
Mean of dep. var. (Placebo)	2.745	0.174	171812.1	928546.1	94136.720	0.171	0.219