

Economics 270B
Ph.D. Development Economics

Professor Ted Miguel
Department of Economics
University of California, Berkeley

Lecture 4 – February 23, 2015

I. Overview of International Economic Development

Lecture 1: Understanding economic growth and development (1/26)

Lecture 1B: Persistence of historical institutions and shocks
(read during holiday week of 2/16)

Lecture 2: The Psychology of Poverty (2/2)

II. Human Capital in Economic Development

Lectures 3-4: Education (2/9, 2/23)

Lectures 5-7: Health and nutrition (3/2, 3/9, 3/16)

III. Political economy

Lectures 8-9: Democracy, Corruption and Development (3/30, 4/6)
(guest lectures by Prof. Fred Finan)

Lecture 10: Ethnic and Social Divisions (4/13)

Lectures 11-12: The Political Economy of Conflict (4/20, 4/27)

I. Overview of International Economic Development

Lecture 1: Understanding economic growth and development (1/26)

Lecture 1B: Persistence of historical institutions and shocks
(read during holiday week of 2/16)

Lecture 2: The Psychology of Poverty (2/2)

II. Human Capital in Economic Development

Lectures 3-4: Education (2/9, 2/23)

Lectures 5-7: Health and nutrition (3/2, 3/9, 3/16)

III. Political economy

Lectures 8-9: Democracy, Corruption and Development (3/30, 4/6)
(guest lectures by Prof. Fred Finan)

Lecture 10: Ethnic and Social Divisions (4/13)

Lectures 11-12: The Political Economy of Conflict (4/20, 4/27)

- Prerequisites: Graduate economic theory, econometrics
- Grading:
 - Four referee reports – 40%
 - Report #2 on Dizon-Ross paper due today (2/23)
 - Report #3 on Morjaria paper due in two weeks (3/9)

Two problem sets – 20%

Research proposal – 30%

Class participation – 10%

No final exam

- All readings are available on bCourses

Any questions?

Lecture 4 outline

- 1) The rise of experimental methods in (development) economics, and other methodological innovations
- 2) Muralidharan and Sundararaman (2011) – teacher performance pay in India
- 3) Baird, McIntosh and Ozler (2011) – conditional versus unconditional cash transfers for adolescents in Malawi
- 4) Kremer, Miguel and Thornton (2009) – student incentives for girls in Kenya

(1) Randomized controlled trials

- As background: randomized controlled trials (**RCTs**) have long been common in medical trials and agricultural research but not in the social sciences
- In economics, “field experiments” first became widespread in development economics starting in 1995 with the education experiments led by Michael Kremer in Kenya, and the Mexico PROGESA experiment in 1997

(1) Randomized controlled trials

- Randomized provision of an education (or other) intervention breaks the link between household characteristics, (unobserved) child innate ability, prior investments in child learning, and education outcomes

(1) Randomized controlled trials


- Randomized provision of an education (or other) intervention breaks the link between household characteristics, (unobserved) child innate ability, prior investments in child learning, and education outcomes
- There may be **endogenous behavioral responses** to an intervention. Thus the difference between the treatment / control groups should be thought of as the combined impact of the intervention *per se* together with any resulting behavioral changes (though these changes can also be measured)
- Different emphasis than in clinical trials in medicine

Randomized experiments in development

Year: 2003

J-PAL + CEGA Evaluations: 40

Affiliates
J-PAL + CEGA = 7

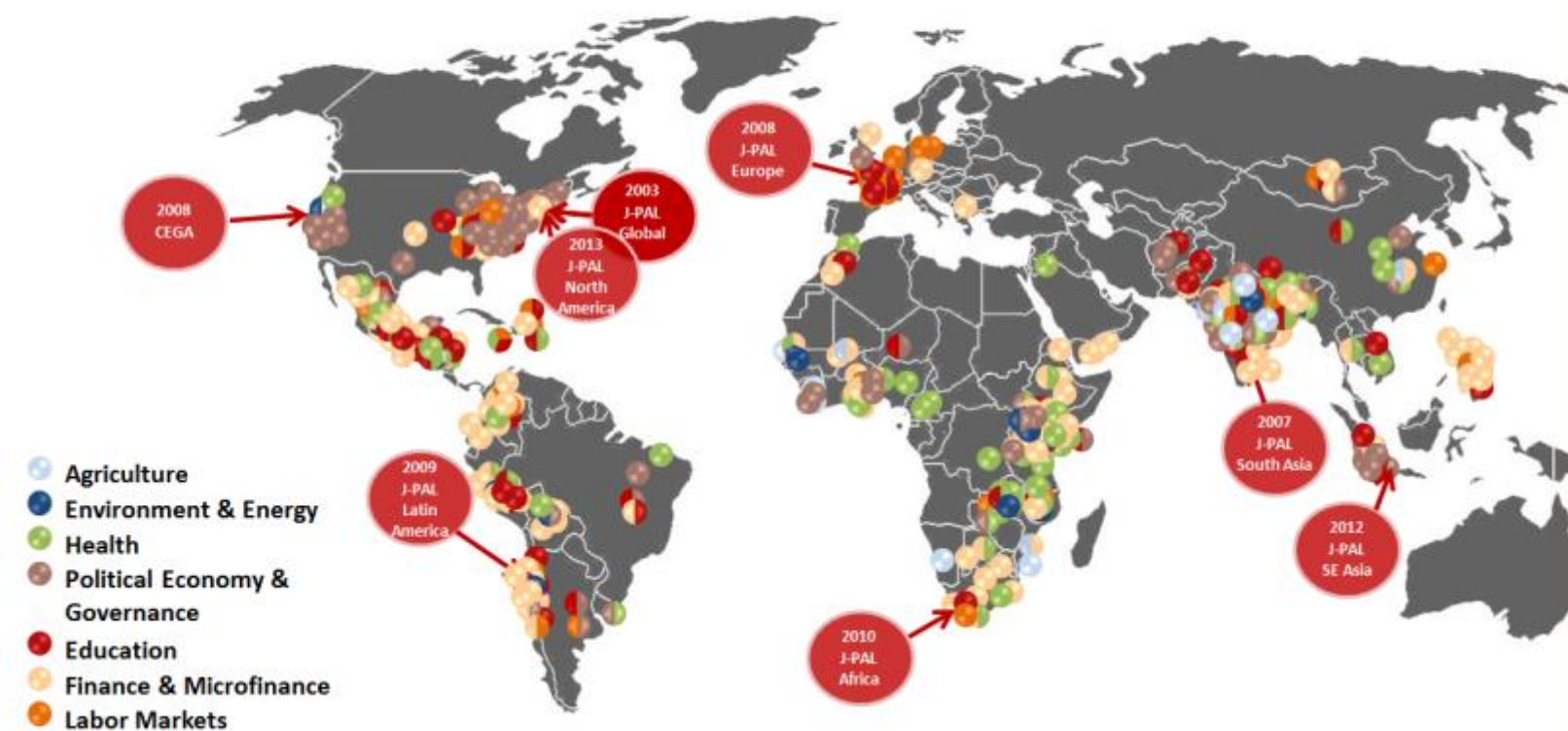


Randomized experiments in development

Year: 2013

J-PAL + CEGA Evaluations: 492

Affiliates
J-PAL + CEGA = 126



Randomized experiments in development

- Randomized experiments have since become common in all applied fields in economics (e.g., labor economics, public economics), and increasingly in other social sciences (especially political science)
- The rise of “real” experiments in economics is one of the most important scientific and methodological innovations in the social sciences over the past few decades.
- Part of a broad intellectual trend as micro-data collection has improved, computing power has become cheaper, and better econometric tools have been developed

(1) Randomized controlled trials

- Yet the trend towards experimental methods – and empirical work in general – in development economics was criticized by some senior leaders in the field (i.e., the Economic and Political Weekly “debate” in 2005 pitting Bardhan, Basu and Mookherjee vs. Banerjee; Deaton 2007 vs. Banerjee, Duflo and Imbens)
- It is worth briefly discussing this “debate”

(1) Randomized controlled trials

- Why randomize?

(1) Randomized controlled trials

- Why randomize?
 - (1) Randomization helps address an array of well-known biases, e.g., it can resolve the selection problem that often plagues treatment effect estimates

(1) Randomized controlled trials

- Why randomize?
 - (1) Randomization helps address an array of well-known biases, e.g., it can resolve the selection problem that often plagues treatment effect estimates
 - (2) As a result, randomized research designs can allow the researchers to identify behavioral parameters that are of theoretical interest, and that are difficult or impossible to estimate using other methods (e.g., estimating social effects)
 - (3) The results of randomized evaluations are typically more transparent and credible to policymakers

(1) Randomized controlled trials

- What are the limitations of randomized methods?

(1) Randomized controlled trials

- What are the limitations of randomized methods?
 - (1) External validity – estimated impacts are “local”
 - (2) They cannot address all problems (e.g., in macro)

(1) Randomized controlled trials

- What are the limitations of randomized methods?
 - (1) External validity – estimated impacts are “local”
But true for all micro-empirical work (e.g., ICRISAT)
 - (2) They cannot address all problems (e.g., in macro)
That is setting the bar too high for any method

(1) Randomized controlled trials

- What are the limitations of randomized methods?
 - (1) External validity – estimated impacts are “local”
 - (2) They cannot address all problems (e.g., in macro)
 - (3) They are “too easy”, anyone can use them
 - (4) These methods are inherently atheoretical
 - (5) They cannot estimate general equilibrium effects

(1) Randomized controlled trials

- What are the limitations of randomized methods?
 - (1) External validity – estimated impacts are “local”
 - (2) They cannot address all problems (e.g., in macro)
 - (3) They are “too easy”, anyone can use them
This is arguably a strength rather than a weakness
 - (4) These methods are inherently atheoretical
Not true: development economists have long used these methods to tackle fundamental theory issues (e.g., Karlan and Zinman 2006 on moral hazard and adverse selection in credit markets)
 - (5) They cannot estimate general equilibrium effects
Large-scale experiments properly designed (cluster randomizations) can estimate spillovers, price effects

(1) Other methodological advances

- There is a related ongoing move toward “research transparency” in Economics and other fields
- (Full disclosure: I am teaching a whole course on these topics this term, Econ 270D)

(1) Other methodological advances

- There is a related ongoing move toward “research transparency” in Economics and other fields
- (Full disclosure: I am teaching a whole course on these topics this term, Econ 270D)
- Miguel et al (2014) lays out three inter-related approaches to begin addressing these problems:
 1. Disclosure (conflicts of interest, intended research design, data collected)
 2. Open data and materials (to allow others to find errors, extend and replicate work)
 3. **Pre-registration of research hypotheses**

(1) Why pre-register?

- First of all, what is **pre-registration**?
- A researcher posts her/his research hypotheses, the data used to test them, and the planned research design (i.e., methodology) in a publicly available **registry**
- There is obviously a wide range of detail one could potentially include in an analysis plan
- Both clinicaltrials.gov and the new AEA registry (discussed next) allow researchers to include relatively sparse information if they choose

(1) Why pre-register?


- The American Economics Association (AEA) registry, **socialscienceregistry.org**, was founded in May 2013 with a focus on randomized control trials (RCTs).

Google News x Google Calendar x Google News x Golden State Files clinicaltrials.gov AEA RCT Registry x how do you print x

← → ↻ 🏠 <https://www.socialscienceregistry.org/site/about> 🔍 ☆ ☰

Apps Google News Google Gmail: Email from G... Google Calendar Other bookmarks

Create Account Sign in

 **AEA RCT Registry**
The American Economic Association's registry for randomized controlled trials

About RCTs Registration Guidelines FAQ

Advanced Search **SEARCH**

ABOUT THE REGISTRY

REGISTER A TRIAL >

Welcome.

This is the American Economic Association's registry for randomized controlled trials.

Randomized Controlled Trials (RCTs) are widely used in various fields of economics and other social sciences. As they become more numerous, a central registry on which trials are on-going or complete (or abandoned) becomes important for various reasons: as a source of results for meta-analysis; as a one-stop resource to find out about available survey instruments and data.

Because existing registries are not well suited to the need for social sciences, in April 2012, the AEA executive committee decided to establish such a registry for economics and other social sciences.

If you are running or have run a trial: Registration is free and you do not need to be a member of the AEA to register. We encourage you to register any new study at its outset. However, given the backlog of existing trials, we invite you to also register past studies.

If you are searching for results: Please browse the data base. More results are forthcoming!

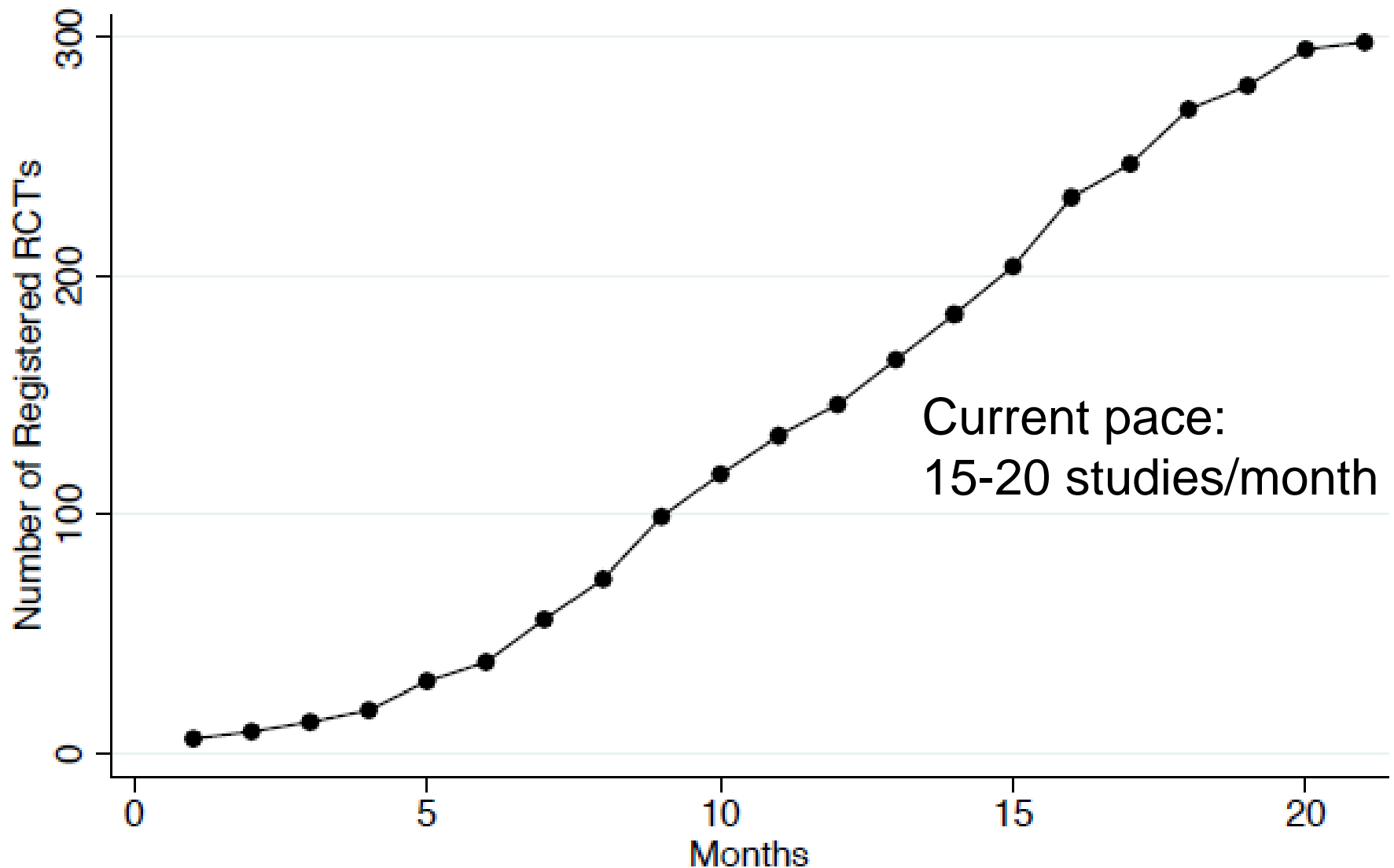
© Copyright 2012-2015, MIT. [About](#) [FAQ](#) [Contact](#)

(1) Why pre-register?

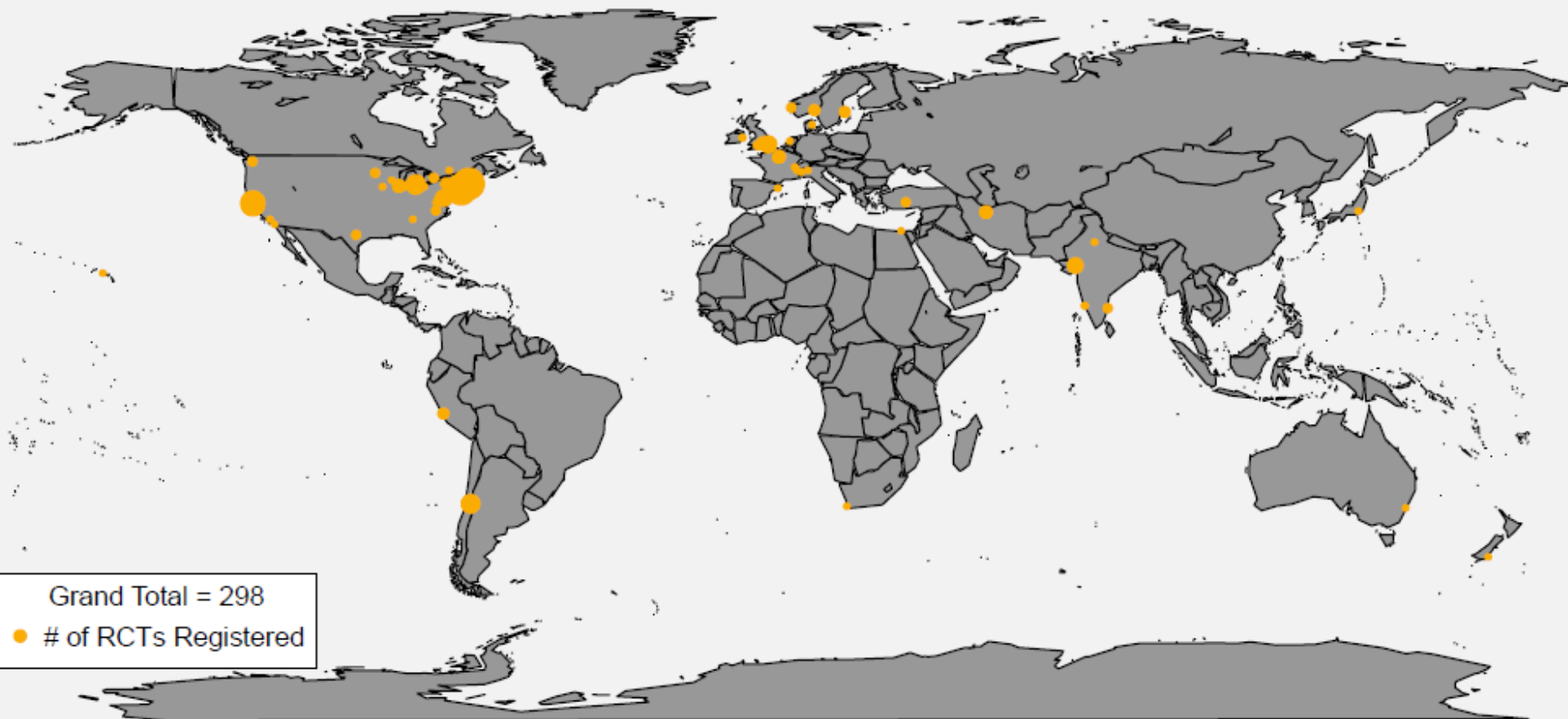
- The American Economics Association (AEA) registry, **socialscienceregistry.org**, was founded in May 2013 with a focus on randomized control trials (RCTs).
- Since then over 300 studies have been registered, and the numbers are increasing rapidly.
- Some of these are earlier projects that are being registered (for completeness), but most are new studies.

Total RCT's Registered in AEA Site

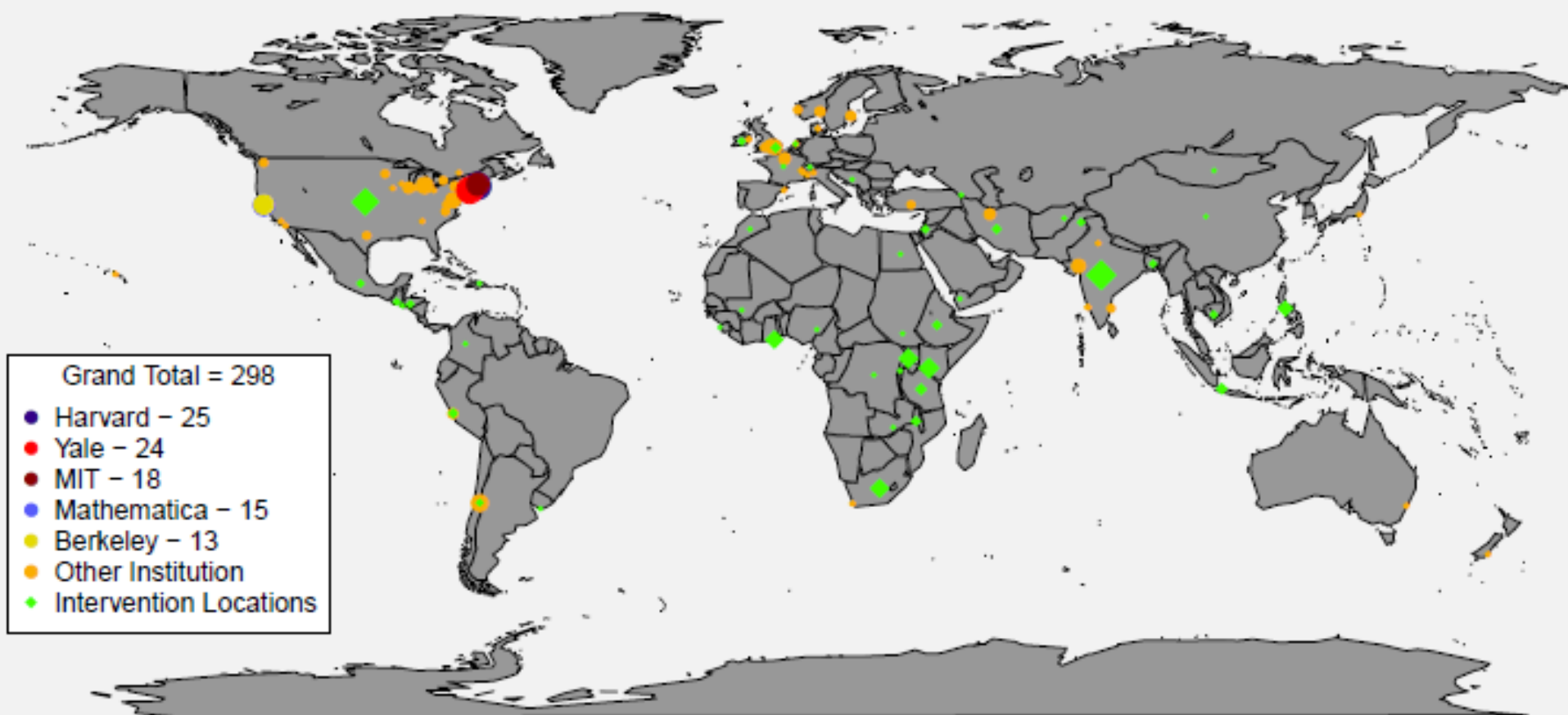
05/2013 - 01/2015



Total RCTs Registered by Research Institution – January 2015



Total RCTs Registered – Jan. 2015



(1) Why pre-register?

- What if your study is not an RCT, but you want to pre-register elements of your analysis?
- The **Open Science Framework** (OSF) provides a flexible platform for time-stamping and archiving materials to be made publicly available
- Or you can post it as a working paper (i.e., NBER WP or CECA WP), in order to time-stamp and archive it.

(1) Why pre-register?

- As I discuss in Econ 270D, there remains some debate about just how detailed the pre-registration material / analysis plan should be, but even with the relatively sparse information on the AEA site, a reader can figure out your main hypotheses, outcome measures, and research design – and that is valuable.
- What concrete benefits could this information have?

(1) Why pre-register?

- **Why might pre-registration be useful?**
 1. **Rounds out the body of evidence** by creating a “paper trail” of unpublished studies in an area → potentially helping to address publication bias (see Franco et al 2014 in *Science*) and improve meta-analysis.

(1) Why pre-register?

- **Why might pre-registration be useful?**
 1. **Rounds out the body of evidence** by creating a “paper trail” of unpublished studies in an area
 2. **Reduces the risk of data mining** and other tendentious presentation of results (“cherry-picking”) → by making clear what the authors’ original intentions and research hypotheses actually were.

(1) Why pre-register?

- **Why might pre-registration be useful?**
 - 1. Rounds out the body of evidence** by creating a “paper trail” of unpublished studies in an area
 - 2. Reduces the risk of data mining** and other tendentious presentation of results (“cherry-picking”)
 - 3. Generates correctly sized statistical tests**, bolstering the credibility of statistical significance levels → by making clear what additional tests were run beyond those originally planned, and thus making multiple testing adjustments more credible.

(1) Why pre-register?

- **Why might pre-registration be useful?**
 - 1. Rounds out the body of evidence** by creating a “paper trail” of unpublished studies in an area
 - 2. Reduces the risk of data mining** and other tendentious presentation of results (“cherry-picking”)
 - 3. Generates correctly sized statistical tests**, bolstering the credibility of statistical significance levels
 - 4. Makes open data and disclosure more effective** → by allowing other scholars to cross-check published information against original research plans.

(1) Why pre-register?

- **Why might pre-registration be useful?**
 1. **Rounds out the body of evidence** by creating a “paper trail” of unpublished studies in an area
 2. **Reduces the risk of data mining** and other tendentious presentation of results (“cherry-picking”)
 3. **Generates correctly sized statistical tests**, bolstering the credibility of statistical significance levels
 4. **Makes open data and disclosure more effective**
 5. As a side benefit, **forces researchers to more carefully think through their hypotheses beforehand, improving research quality** → reducing “waste” of funding on poorly conceived projects

(1) Why pre-register?

- **Why might pre-registration be useful?**
 1. **Rounds out the body of evidence** by creating a “paper trail” of unpublished studies in an area
 2. **Reduces the risk of data mining** and other tendentious presentation of results (“cherry-picking”)
 3. **Generates correctly sized statistical tests**, bolstering the credibility of statistical significance levels
 4. **Makes open data and disclosure more effective**
 5. As a side benefit, **forces researchers to more carefully think through their hypotheses beforehand**
 6. **Others?**

(1) Why pre-register?

- A leading concern: will pre-registration of plans **stifle creativity** and limit discoveries made through exploratory research?
- Many, if not most, important scientific findings undoubtedly originated as unexpected discoveries...

(1) Why pre-register?

- A leading concern: will pre-registration of plans **stifle creativity** and limit discoveries made through exploratory research?
 - Many, if not most, important scientific findings undoubtedly originated as unexpected discoveries...
 - But findings from such work are inherently more tentative because of the greater flexibility of tests and the greater opportunity for the outcome to obtain by chance.
- Pre-specification is not intended to disparage exploratory analysis, but rather to **free it from the tradition of being portrayed as formal hypothesis testing.**

(1) Why pre-register?

- Pre-registration also shifts some of the **time “costs” of research up front** (rather than after data has been collected), which makes it more challenging to very quickly launch an experiment if a new opportunity arises
- Other potential costs of pre-registration?

(2) Muralidharan and Sundararaman (2011)

(2) Muralidharan and Sundararaman (2011)

- A randomized control trial (RCT) conducted in India
- This study is an exemplar of clear (and ambitious) research design, high-quality data, and policy relevance
- (This work, and other research by the authors, are having major policy impact in India)
- Carried out before pre-registration was widespread in Economics (see Casey et al 2012 *QJE* from the syllabus and Finkelstein et al 2012 *QJE* for two of the earliest economics studies with pre-analysis plans)

(2) Muralidharan and Sundararaman (2011)

- How does the effectiveness of spending on **inputs** compare to improved **incentives**?
- Fundamental education policy issue.
- Will high-powered incentives distort teacher behavior in negative ways? E.g., teaching to the test.

(2) Muralidharan and Sundararaman (2011)

- How does the effectiveness of spending on **inputs** compare to improved **incentives**?
- Fundamental education policy issue.
- Will high-powered incentives distort teacher behavior in negative ways? E.g., teaching to the test.
- Examine a large-scale **randomized evaluation in Andhra Pradesh state**: 300 primary schools
 - Schools tend to be small (3 classrooms / school)
 - Teacher incentives vs. more inputs (same monetary value) provided to random subsets of schools
 - Individual incentives versus group incentives

(2) Muralidharan and Sundararaman (2011)

- Why teacher incentives in India?
 - **25% teacher absenteeism** on any given day
 - Teacher salaries 90% of non-capital education spending
 - Strong teacher unions, no discipline for poor outcomes
- Concerns about incentive programs: cheating, teaching to the test (rather than “real” learning), teacher transfers between schools, political backlash from teachers?

(2) Muralidharan and Sundararaman (2011)

- Punchline: **incentives** matter more than inputs in AP
 - 0.22 s.d. gain in incentive schools (equivalent to 9 percentage points at the median), 0.08 in input schools
 - Math gains (0.27) larger than language gains (0.17)
 - All students at least weakly gain (at least 0.1 s.d.)

(2) Muralidharan and Sundararaman (2011)

- Punchline: incentives matter more than inputs in AP
 - 0.22 s.d. gain in incentive schools (equivalent to 9 percentage points at the median), 0.08 in input schools
 - Math gains (0.27) larger than language gains (0.17)
 - All students at least weakly gain (at least 0.1 s.d.)
 - No real evidence of diversion of efforts away from other subjects, cheating, mechanical / rote learning, or teacher opposition. In fact, positive spillovers with gains in non-incentivized subjects (science, social studies)
- **Mixed results** in other studies: Lavy (2002, 2009) finds strong positive impacts of teacher incentives in Israel, others show weaker impacts in Kenya, U.S.

(2) Muralidharan and Sundararaman (2011)

- Incentives and teacher effort: multitask moral hazard
- Teachers can engage in two types of behaviors, T_1 (“best practice”, time t_1) and T_2 (teaching to the test, time t_2)
- Human capital production: $H = f_1 t_1 + f_2 t_2 + \varepsilon$
 - where f denotes marginal products, ε denotes factors outside of teacher control (e.g., parents)

$$f_1 > f_2$$

(2) Muralidharan and Sundararaman (2011)

- Incentives and teacher effort: multitask moral hazard
- Teachers can engage in two types of behaviors, T_1 (“best practice”, time t_1) and T_2 (teaching to the test, time t_2)
- Human capital production: $H = f_1 t_1 + f_2 t_2 + \varepsilon$
 - where f denotes marginal products, ε denotes factors outside of teacher control (e.g., parents)
- The social planner cannot observe H , rather an imperfect proxy in test scores: $P = g_1 t_1 + g_2 t_2 + \phi$
- Education experts claim: $f_1 > f_2$ and $g_2 > g_1$

(2) Muralidharan and Sundararaman (2011)

- Incentives can only be conditioned on the test score, P : let the teacher's wage $w = \text{salary} + \text{bonus} * P$
- Following Holmstrom and Milgrom (1991), the teacher's utility is given by: $U = E(w) - C(t_1, t_2; t^*)$
- Assume there is a large psychic or social cost if total teacher effort $t_1 + t_2$ falls below the effort "norm" t^*

(2) Muralidharan and Sundararaman (2011)

- Incentives can only be conditioned on the test score, P : let the teacher's wage $w = \text{salary} + \text{bonus} * P$
- Following Holmstrom and Milgrom (1991), the teacher's utility is given by: $U = E(w) - C(t_1, t_2; t^*)$
- Assume there is a large psychic or social cost if total teacher effort $t_1 + t_2$ falls below the effort "norm" t^*
- **In high t^* settings**, $b=0$ could be optimal (to prevent teachers from teaching to the test, thus reducing H)
- **But in low t^* setting** (like India), the gains to increasing total effort could swamp this distortion – especially if f_1/f_2 is close to 1 (e.g., given emphasis already placed on exam preparation in India and other Asian countries)

Figure 1a: Andhra Pradesh (AP)



	India	AP
Gross Enrollment (Ages 6-11) (%)	95.9	95.3
Literacy (%)	64.8	60.5
Teacher Absence (%)	25.2	25.3
Infant Mortality (per 1000)	63	62

3.2 Sampling

We sampled 5 districts across each of the 3 socio-cultural regions of AP in proportion to population (Figure 1b).²² In each of the 5 districts, we randomly selected one division and then randomly sampled 10 mandals in the selected division. In each of the 50 mandals, we randomly sampled 10 schools using probability proportional to enrollment.

Thus, the universe of 500 schools in the study was representative of the schooling conditions of the typical child attending a government-run primary school in rural AP.

3.3 AP RESt Design Overview

The overall design of AP RESt is represented in the table below:

Table 3.1

	INCENTIVES (Conditional on Improvement in Student Learning)			
INPUTS (Unconditional)		NONE	GROUP BONUS	INDIVIDUAL BONUS
	NONE	CONTROL (100 Schools)	100 Schools	100 Schools
	EXTRA PARA TEACHER	100 Schools		
	EXTRA GRANT	100 Schools		

D. Description of Incentive Treatments

Teachers in incentive schools were offered bonus payments on the basis of the average improvement in test scores (in math and language) of students taught by them subject to a minimum improvement of 5 percent. The bonus formula was

$$\text{Bonus} = \begin{cases} \text{Rs. } 500 \times (\% \text{ gain in average test scores} - 5\%) & \text{if gain} > 5\% \\ 0 & \text{otherwise.} \end{cases}$$

All teachers in group incentive schools received the same bonus based on average school-level improvement in test scores, whereas the bonus for teachers in individual incentive schools was based on the average test score improvement of students taught by the specific teacher.¹³ We

TABLE 2
SAMPLE BALANCE ACROSS TREATMENTS

	Control (1)	Group Incentive (2)	Individual Incentive (3)	<i>p</i> -Value (Equality of All Groups) (4)
A. Means of Baseline Variables				
School-level variables:				
1. Total enrollment (baseline: grades 1–5)	113.2	111.3	112.6	.82
2. Total test takers (baseline: grades 2–5)	64.9	62.0	66.5	.89
3. Number of teachers	3.07	3.12	3.14	.58
4. Pupil-teacher ratio	39.5	40.6	37.5	.66
5. Infrastructure index (0–6)	3.19	3.14	3.26	.84
6. Proximity to facilities index (8–24)	14.65	14.66	14.72	.98
Baseline test performance:				
7. Math (raw %)	18.5	18.0	17.5	.69
8. Math (normalized; in SD)	.032	.001	–.032	.70
9. Telugu (raw %)	35.1	34.9	33.5	.52
10. Telugu (normalized; in SD)	.026	.021	–.046	.53

B. Specification

We first discuss the impact of the incentive program as a whole by pooling the group and individual incentive schools and considering this to be the “incentive” treatment. All estimation and inference are done with the sample of 300 control and incentive schools unless stated otherwise. Our default specification uses the form

$$T_{ijkm}(Y_n) = \alpha + \gamma \cdot T_{ijkm}(Y_0) + \delta \cdot \text{Incentives} + \beta \cdot Z_m + \varepsilon_k + \varepsilon_{jk} + \varepsilon_{ijk} \quad (1)$$

The main dependent variable of interest is T_{ijkm} , which is the normalized test score on the specific subject, where i , j , k , and m denote the student, grade, school, and mandal, respectively. The term Y_0 indicates the baseline tests, and Y_n indicates a test at the end of n years of the program. Including the normalized baseline test score improves efficiency as a result of the autocorrelation between test scores across multiple periods.²⁴ All regressions include a set of mandal-level dummies (Z_m), and the standard errors are clustered at the school level. We also run the regressions with and without controls for household and school variables. The Incentives variable is a dummy at the school level indicating treatment status, and the parameter of interest is δ , which is the effect on test scores of being in an incentive school. The random assignment of the incentive program ensures that this is an unbiased and consistent estimate of the 1-year and 2-year treatment effects.

TABLE 3
IMPACT OF INCENTIVES ON STUDENT TEST SCORES
Dependent Variable: Normalized End-of-Year Test Score

	YEAR 1 ON YEAR 0		YEAR 2 ON YEAR 0	
	(1)	(2)	(3)	(4)
A. Combined (Math and Language)				
Normalized lagged test score	.503*** (.013)	.498*** (.013)	.452*** (.015)	.446*** (.015)
Incentive school	.149*** (.042)	.165*** (.042)	.219*** (.047)	.224*** (.048)
School and household controls	No	Yes	No	Yes
Observations	42,145	37,617	29,760	24,665
R^2	.31	.34	.24	.28
B. Math				
Normalized lagged test score	.492*** (.016)	.491*** (.016)	.414*** (.022)	.408*** (.022)
Incentive school	.180*** (.049)	.196*** (.049)	.273*** (.055)	.280*** (.056)
School and household controls	No	Yes	No	Yes
Observations	20,946	18,700	14,797	12,255
R^2	.30	.33	.25	.28

(2) Muralidharan and Sundararaman (2011)

- Similar effects on repeat and new questions; conceptual and rote questions; multiple choice and written questions – suggesting real learning (Tables 4, 5)
- Gains throughout the test score distribution, and significant gains above roughly 40th percentile (figure 2)
- Heterogeneous treatment effects: few meaningful interactions with household, school or student characteristics, including baseline test score, but classes with more educated (experienced) teachers show larger (smaller) impacts (Table 6)
- Provides some guidance for extent of external validity

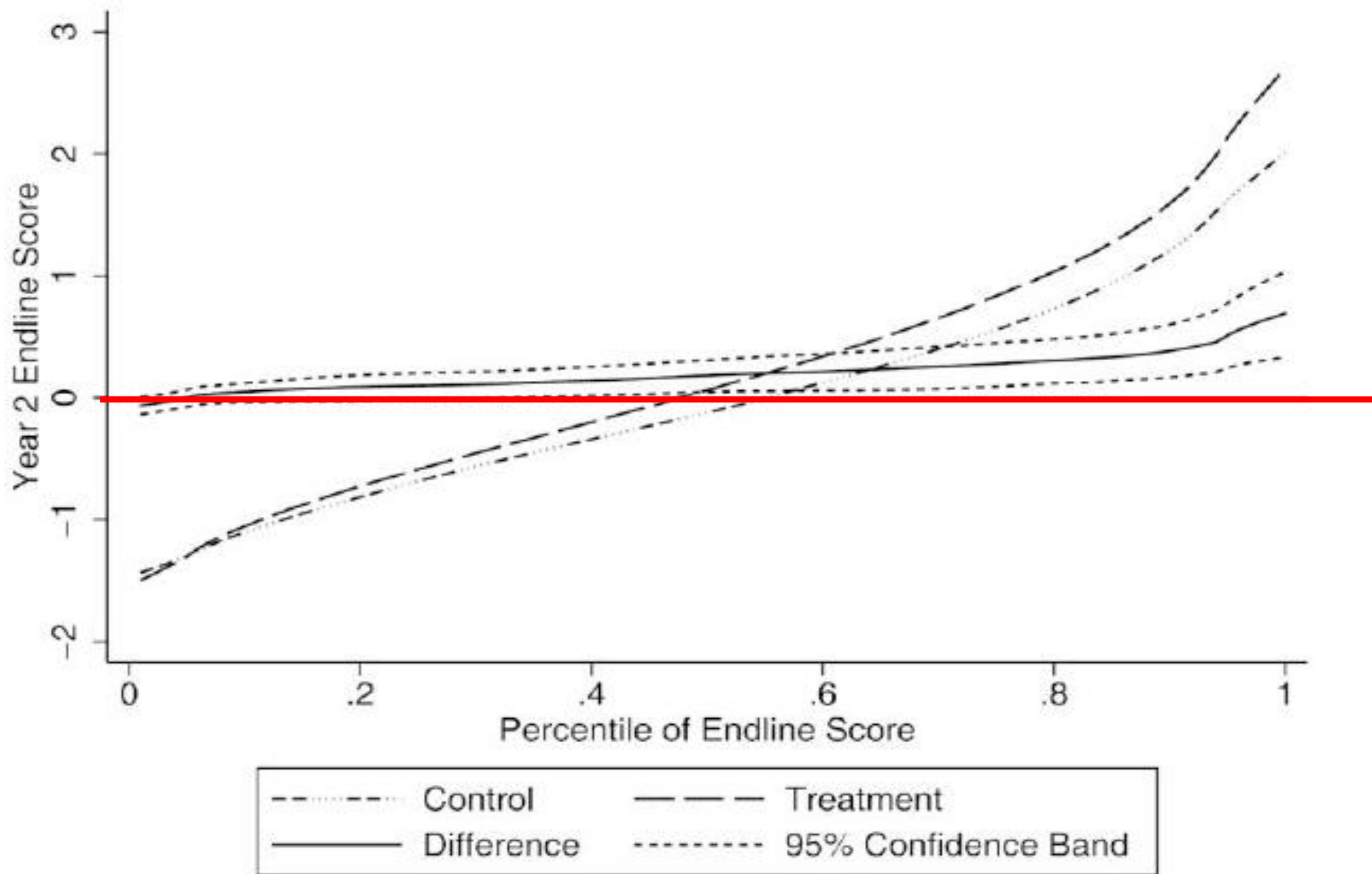


FIG. 2.—Quantile treatment effects of the performance pay program on student test scores.

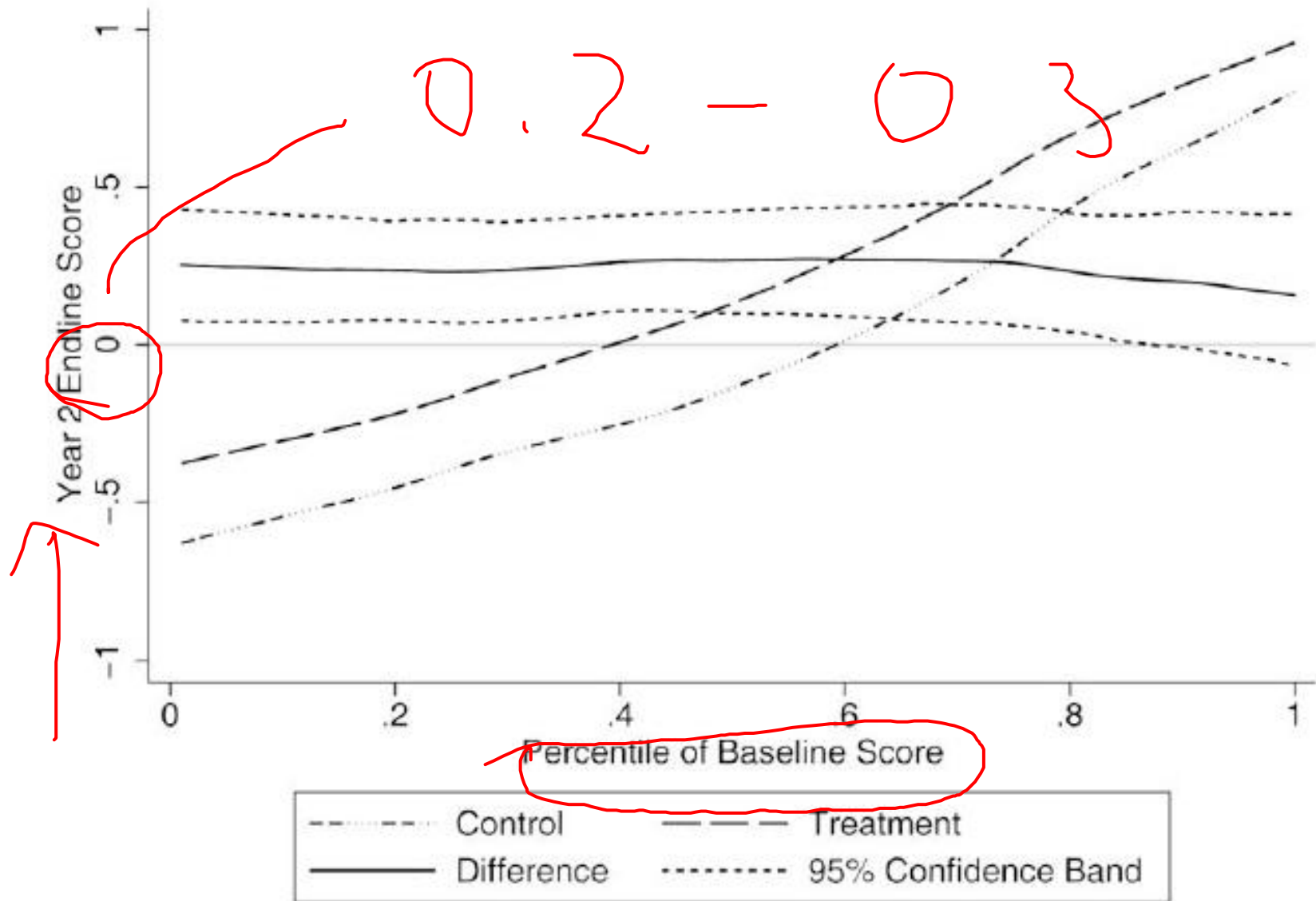


FIG. 3.—Nonparametric treatment effects by percentile of baseline score

TABLE 6
HETEROGENOUS TREATMENT EFFECTS
A. HOUSEHOLD AND SCHOOL CHARACTERISTICS

	Log School Enrollment (1)	School Proximity (8–24) (2)	School Infrastructure (0–6) (3)	Household Affluence (0–7) (4)	Parental Literacy (0–4) (5)	Scheduled Caste/Tribe (6)	Male (7)	Normalized Baseline Score (8)
Two-Year Effect								
Incentive	–.198 (.354)	–.019 (.199)	.28** (.130)	.09 (.073)	.224*** (.054)	.226*** (.049)	.233*** (.049)	.219*** (.047)
Covariate	–.065 (.058)	–.005 (.010)	.025 (.038)	.017 (.014)	.068*** (.015)	–.066 (.042)	.029 (.027)	.448*** (.024)
Interaction	.083 (.074)	.018 (.014)	–.02 (.040)	.038** (.019)	–.003 (.019)	–.013 (.056)	–.02 (.034)	.006 (.031)
Observations	29,760	29,760	29,760	25,231	25,226	29,760	25,881	29,760
R^2	.244	.244	.243	.272	.273	.244	.266	.243

B. TEACHER CHARACTERISTICS (Pooled Regression Using Both Years of Data)

	Education (1)	Training (2)	Years of Experience (3)	Salary (Log) (4)	Male (5)	Teacher Absence (6)	Active Teaching (7)	Active or Passive Teaching (8)
Incentive	–.113 (.163)	–.224 (.176)	.258*** (.059)	1.775** (.828)	.031 (.091)	.15*** (.050)	.084 (.054)	.118 (.074)
Covariate	.003 (.032)	–.051 (.041)	–.001 (.003)	–.034 (.066)	–.084 (.057)	–.149 (.137)	.055 (.078)	.131 (.093)
Interaction	.086* (.050)	.138** (.061)	–.009** (.004)	–.179* (.091)	.09 (.069)	.013 (.171)	.164* (.098)	.064 (.111)
Observations	53,737	53,890	54,142	53,122	54,142	53,609	53,383	53,383
R^2	.29	.29	.29	.29	.29	.29	.29	.29

(2) Muralidharan and Sundararaman (2011)

- Spillovers of similar magnitude to other subjects, probably due to language and math gains (Table 7)
- No difference between individual and group incentives. Recall these are **very small schools**, with 80 students and 3 teachers on average. (How would results generalize to larger schools, e.g., with 10 teachers?)
- Perhaps surprisingly, no difference in student, teacher attendance between the treatment and control groups – but big differences in test preparation (Table 9)

TABLE 7
IMPACT OF INCENTIVES ON NONINCENTIVE SUBJECTS
Dependent Variable: Normalized End Line Score

	YEAR 1		YEAR 2	
	Science	Social Studies	Science	Social Studies
A. Reduced-Form Impact				
Normalized baseline math score	.215*** (.019)	.224*** (.018)	.156*** (.023)	.167*** (.024)
Normalized baseline language score	.209*** (.019)	.289*** (.019)	.212*** (.023)	.189*** (.024)
Incentive school	.112** (.052)	.141*** (.048)	.113** (.044)	.18*** (.050)
Observations	11,786	11,786	9,143	9,143
R^2	.26	.31	.19	.18

TABLE 9
TEACHER BEHAVIOR (Observation and Interviews)

TEACHER BEHAVIOR	INCENTIVE VERSUS CONTROL SCHOOLS (%)			Correlation with Student Test Score Gains (4)
	Incentive Schools (1)	Control Schools (2)	<i>p</i> -Value of Difference (3)	
Teacher absence (%)	.25	.23	.199	-.103
Actively teaching at point of ob- servation (%)	.42	.43	.391	.135***
Did you do any special prepara- tion for the end of year tests? (% Yes)	.64	.32	.000***	.095**
What kind of preparation did you do? (unprompted; % mentioning):				
Extra homework	.42	.20	.000***	.061
Extra classwork	.47	.23	.000***	.084**
Extra classes/teaching be- yond school hours	.16	.05	.000***	.198***
Gave practice tests	.30	.14	.000***	.105**
Paid special attention to weaker children	.29	.07	.000***	.010

(2) Muralidharan and Sundararaman (2011)

- Further questions / issues:
 - Back to the original question: what is the relative impact of funding spent on incentives versus inputs?

TABLE 10
IMPACT OF INPUTS VERSUS INCENTIVES ON LEARNING OUTCOMES
Dependent Variable: Normalized End-of-Year Test Score

	YEAR 1 ON YEAR 0			YEAR 2 ON YEAR 0		
	Combined (1)	Math (2)	Language (3)	Combined (4)	Math (5)	Language (6)
Normalized lagged score	.512*** (.010)	.494*** (.012)	.536*** (.011)	.458*** (.012)	.416*** (.016)	.499*** (.012)
Incentives	.15*** (.041)	.179*** (.048)	.121*** (.039)	.218*** (.049)	.272*** (.057)	.164*** (.046)
Inputs	.102*** (.038)	.117*** (.042)	.086** (.037)	.085* (.046)	.089* (.052)	.08* (.044)
<i>F</i> -statistic <i>p</i> -value (inputs = incen- tives)	.178	.135	.298	.003	.000	.044
Observations	69,157	34,376	34,781	49,503	24,628	24,875
R^2	.30	.29	.32	.225	.226	.239

(2) Muralidharan and Sundararaman (2011)

- Potentially a very **high rate of return**:
 - Cross-sectional estimates from India suggest a return of roughly 16% (20%) for scoring 1 SD higher on a standardized math (language) test
 - If these are additive, the two year effect of this program implies an increase of roughly $(0.27 \text{ SD} \times 16\%/\text{SD} + 0.17 \text{ SD} \times 20\%/\text{SD}) = \mathbf{7.7\%}$ increase in wages
 - Extremely high internal rate of return (discounted future gains at least 16x larger than costs)

(2) Muralidharan and Sundararaman (2011)

- Further questions / issues:
 - Back to the original question: what is the relative impact of funding spent on incentives versus inputs?
 - What is the optimal incentive contract? How steep should incentives be? How large are the utility costs for risk averse teachers?
 - Would **steeper incentives** (eventually) lead more talented individuals into the teaching profession?
 - Would teacher unions allow these experiments on a wider scale? Any political backlash?
 - How to boost teacher “value added” more generally beyond incentives? Status, work conditions, technology?

(2) Muralidharan and Sundararaman (2011)

- Any additional comments?

(3) Baird et al (2011)

- Incentives appear to improve teacher effort and performance – but what about students?

(3) Baird et al (2011)

- Incentives appear to improve teacher effort and performance – but what about students?
- The most common incentive in large-scale programs today are conditional cash transfer (**CCT**) programs, many modeled on the Mexico Progresa program – which showed large impacts in a randomized evaluation (as have studies in Ecuador, Brazil, etc.).
- By 2007, 29 countries had a CCT program (World Bank 2009), and many more since then. Unconditional cash transfer programs (UCT) are also common (South Africa, Uruguay).
- Political advantages of CCT? Logistical disadvantages?

(3) Baird et al (2011)

- A randomized evaluation took place in **161 enumeration areas** (EA's) in Zomba, Malawi over two years (2008-2009). $N_{\text{CCT}} = 46$, $N_{\text{UCT}} = 27$.
- Beneficiaries: adolescent girls ($N=2,284$) and parents each got randomly determined transfer amounts (\$1-5 and \$4-10, respectively), and had school fees paid.

(3) Baird et al (2011)

- A randomized evaluation took place in **161 enumeration areas** (EA's) in Zomba, Malawi over two years (2008-2009). $N_{\text{CCT}} = 46$, $N_{\text{UCT}} = 27$.
- Beneficiaries: adolescent girls ($N=2,284$) and parents each got randomly determined transfer amounts (\$1-5 and \$4-10, respectively), and had school fees paid.
- In the CCT arm, transfers were conditioned on school attendance ($\geq 80\%$, ≤ 4 absences/month). Do conditions help youth “bargain” with parents over schooling?
- UCT's are logistically easier (and cheaper) to administer. Are the conditions essential or can they be dropped?

(3) Baird et al (2011)

- Program eligibility for **never married females 13-22 years old**, for both school drop-outs and those enrolled in school. Study focuses on baseline schoolgirls.

(3) Baird et al (2011)

- Program eligibility for **never married females 13-22 years old**, for both school drop-outs and those enrolled in school. Study focuses on baseline schoolgirls.
- Study elements:
 - 1) Rich original data collection:
 - 2) Separate household and participant surveys (including on relationships, sexual behavior, **teen pregnancy**)
 - 3) School enrollment, attendance from registers (without spot checks, to avoid misunderstandings in UCT areas)
 - 4) Endline math, English, cognitive (Ravens) tests
 - 5) School surveys; structured interviews at endline
 - 6) HIV and STI biomarker data (follow-up *Lancet* paper)

(3) Baird et al (2011)

- Research design checks:
- Balanced treatment and control groups at baseline
- Minimal attrition over time (5% after one year, <10% after two years), but not perfectly balanced across groups, with higher tracking rates in the treatment arms
- School enrollment self-reports are more reliable in the CCT group than in either UCT or control, presumably because they knew their enrollment was being monitored

(3) Baird et al (2011)

- This paper is the first to rigorously assess the impacts of the conditions. A fascinating constellation of results:
- Larger **education gains** (enrollment, attendance, tests) in the CCT group than in UCT. I.e., School enrollment rises by 0.18 years in CCT, 0.08 UCT
- **But larger drops in marriage (44%) and fertility (27%)** among adolescent girls in the UCT group than in CCT.
- Puzzle? CCT has income effects and incentives (both of which should reduce adolescents' marriage, fertility), while UCT just has income effects. Cash transfers could help young women avoid “sugar daddies” (Dupas 2011)

TABLE V
PROGRAM IMPACTS ON ATTENDANCE FROM SCHOOL LEDGERS

	Dependent variable: Fraction of days respondent attended school				
	(1)	(2)	(3)	(4)	(5)
	Term 1, 2009	Term 2, 2009	Term 3, 2009	Overall 2009	Term 1, 2010
Conditional treatment	0.139*** (0.045)	0.014 (0.033)	0.169** (0.085)	0.080** (0.035)	0.092** (0.041)
Unconditional treatment	0.063 (0.056)	0.038 (0.033)	0.118 (0.102)	0.058 (0.037)	-0.038 (0.053)
Mean in the control group	0.778	0.849	0.688	0.810	0.801
Number of observations	284	285	192	319	211
Prob > F (Conditional = Unconditional)	0.129	0.334	0.358	0.436	0.010

TABLE VI
PROGRAM IMPACTS ON TEST SCORES

	Dependent variable			
	(1)	(2)	(3)	(4)
	English test score (standardized)	TIMMS math score (standardized)	Non-TIMMS math score (standardized)	Cognitive test score (standardized)
Conditional treatment	0.140*** (0.054)	0.120* (0.067)	0.086 (0.057)	0.174*** (0.048)
Unconditional treatment	-0.030 (0.084)	0.006 (0.098)	0.063 (0.087)	0.136 (0.119)
Number of observations	2,057	2,057	2,057	2,057
Prob > F (Conditional=Unconditional)	0.069	0.276	0.797	0.756

TABLE VII
PROGRAM IMPACTS ON MARRIAGE AND PREGNANCY

	Dependent variable			
	(1)	(2)	(3)	(4)
	=1 if ever married		=1 if ever pregnant	
	Round 2	Round 3	Round 2	Round 3
Conditional treatment	0.007 (0.012)	-0.012 (0.024)	0.013 (0.014)	0.029 (0.027)
Unconditional treatment	-0.026** (0.012)	-0.079*** (0.022)	-0.009 (0.017)	-0.067*** (0.024)
Mean in the control group	0.043	0.180	0.089	0.247
Number of observations	2,087	2,084	2,086	2,087
Prob > F (Conditional = Unconditional)	0.024	0.025	0.265	0.003

(3) Baird et al (2011)

- Puzzling set of findings, especially given existing African evidence that reducing school dropout should lead to declines in teen marriage and pregnancy (Duflo, Dupas, and Kremer 2010; Ozier 2010; Ferré 2009, etc.).

(3) Baird et al (2011)

- Puzzling set of findings, especially given existing African evidence that reducing school dropout should lead to declines in teen marriage and pregnancy (Duflo, Dupas, and Kremer 2010; Ozier 2010; Ferré 2009, etc.).
- A simple explanation can make sense of the patterns. Imagine there are three groups of girls:
 - 1) “UCT compliers”: attend school if receive UCT
 - 2) “CCT compliers”: attend school if receive CCT (with additional conditionality) but not if receive UCT
 - 3) “Non-compliers”: do not attend with CCT or UCT

(3) Baird et al (2011)

- Puzzling set of findings, especially given existing African evidence that reducing school dropout should lead to declines in teen marriage and pregnancy (Duflo, Dupas, and Kremer 2010; Ozier 2010; Ferré 2009, etc.).
- A simple explanation can make sense of the patterns. Imagine there are three groups of girls:
 - 1) “UCT compliers”: attend school if receive UCT [61%]
 - 2) “CCT compliers”: attend school if receive CCT (with additional conditionality) but not if receive UCT [8%]
 - 3) “Non-compliers”: do not attend with CCT or UCT [31%]

TABLE VIII
PREVALENCE OF BEING EVER MARRIED BY SCHOOL ENROLLMENT STATUS DURING
TERM 1, 2010

	(1)	(2)	(3)
	Enrolled	Not enrolled	Total
Control, % (row %)	1.7 (59.8)	46.9 (40.2)	19.9 (100.0)
Conditional treatment, % (row %)	0.5 (69.2)	50.8 (30.8)	16.0 (100.0)
Unconditional treatment, % (row %)	0.3 (60.5)	25.2 (39.5)	10.1 (100.0)
Total, % (row %)	1.1 (62.7)	44.2 (37.3)	17.2 (100.0)

Notes. This table presents the marriage rates by Round 3 enrollment status in Term 1, 2010 and treatment status. For each treatment arm, the top row summarizes the marriage rates by follow-up enrollment status, and the bottom row shows the raw follow-up row percentage in each cell. Means are weighted to make them representative of the target population in the study EAs.

TABLE IX

TEACHER-REPORTED SCHOOL ENROLLMENT AND MARITAL STATUS IN ROUND 3

	Dependent variable			
	(1)	(2)	(3)	(4)
	=1 if enrolled term 1 2010	=1 if ever married	=1 if ever married	=1 if ever married
	All	All	Enrolled	Not enrolled
Conditional treatment	0.058* (0.034)	−0.026 (0.037)	−0.012 (0.015)	0.033 (0.097)
Unconditional treatment	−0.000 (0.036)	−0.088*** (0.030)	−0.011 (0.010)	−0.159** (0.067)
Mean in the control group	0.598	0.199	0.017	0.469
Sample size	844	844	490	354
Prob > F (Conditional = Unconditional)	0.099	0.106	0.857	0.088

(3) Baird et al (2011)

- Out-of-school adolescent girls in Malawi (and other African settings) have poor labor market prospects, and high rates of marriage and pregnancy.
- The results of this study indicate that relatively small amounts of income can boost their “autonomy” and ability to resist pressures (family, social) to get married.
- A downside of conditional cash transfers here is that **the most vulnerable young women** (who would drop out anyway, due to poor academic performance or household poverty) receive less income than under UCT

(3) Baird et al (2011)

- This introduces another factor policymakers need to consider when designing social programs. How large are the three “strata” likely to be? **Key trade-off**: educational gains among the “CCT compliers” versus lower income for the “Non-compliers”.

(3) Baird et al (2011)

- This introduces another factor policymakers need to consider when designing social programs. How large are the three “strata” likely to be? **Key trade-off**: educational gains among the “CCT compliers” versus lower income for the “Non-compliers”.
- Context specific? Malawi is among the poorest African countries: per capita GDP (PPP) US\$760. Different policy choices if fewer “non-compliers” elsewhere?
- The authors also find significant **reductions (by more than half) in HIV and STI infections** among CCT/UCT girls (pooled) relative to the control group by 2010 → non-public health interventions can have major impacts on the spread of HIV

(4) Kremer, Miguel, Thornton (2009)

- A different example of incentives for students: merit scholarships and schooling in rural Kenya
- The debate over merit scholarships
 - “Pros”: Incentives to exert effort, perhaps helping to deal with self-control problems or externalities to effort
 - Possible “cons”:
 - (1) Exacerbate inequality
 - (2) **Weaken intrinsic motivation** in short or long run
 - (3) Gaming the system through cramming, cheating, less effort in other key dimensions

The Girls Scholarship Program (GSP)

- GSP is a randomized evaluation of a merit award for Grade 6 girls in Busia and Teso districts, Kenya
- 64 treatment schools, 63 comparison schools
- The top 15% of girls in program schools (by district) received a \$38 prize for school fees and supplies over two years, and a public awards ceremony

Two GSP research questions

(#1) What impact do these incentives have on test scores and other measures of school performance?

→ Randomized evaluation methods

(#2) What impact does winning the GSP award have on later schooling choices and outcomes? In particular does it make it more likely that winners stay in school?

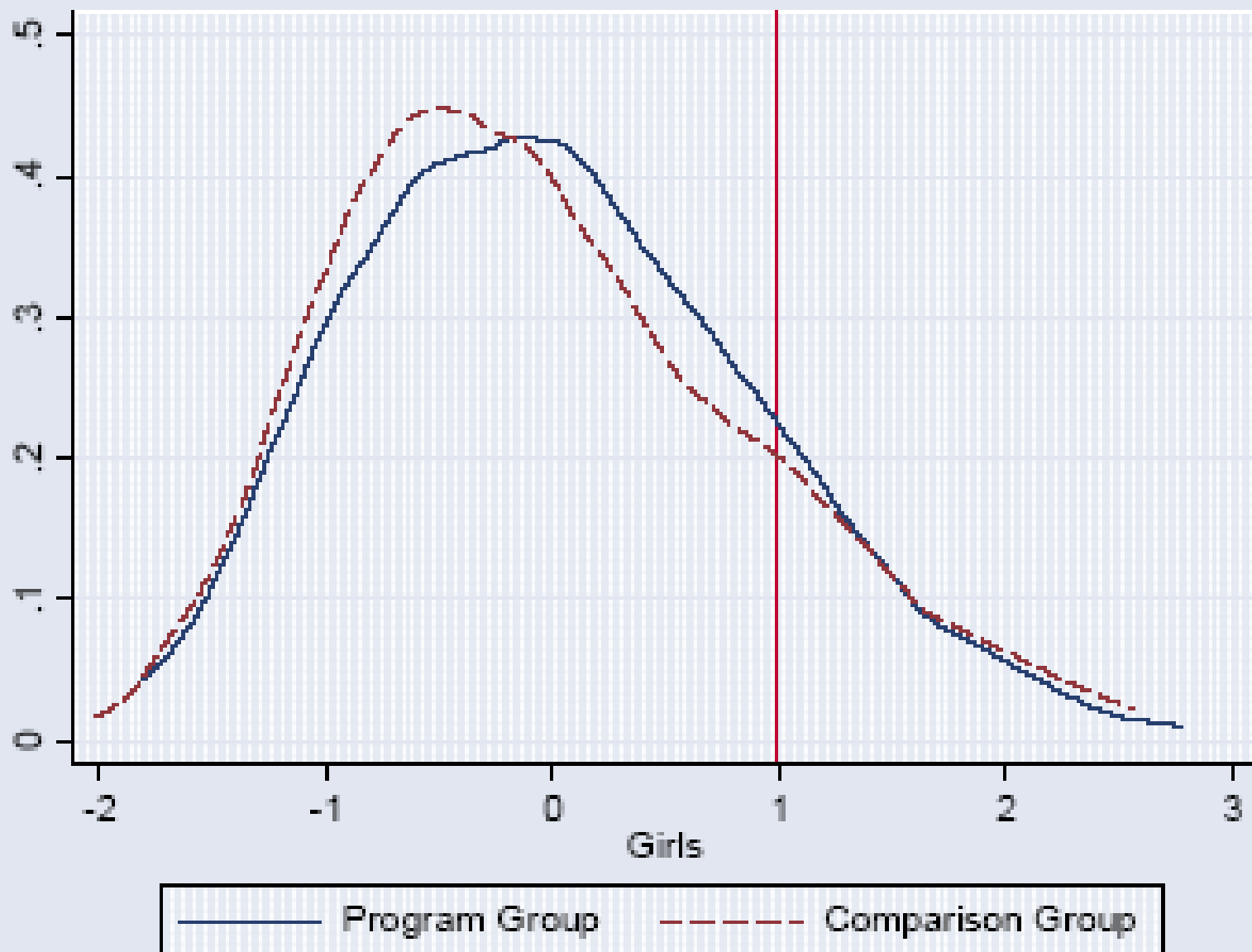
→ Regression discontinuity methods

The Girls Scholarship Program (GSP)

- The randomization “worked”: treatment and comparison group schools are similar at baseline (Table 3, Figure 5)

Panel A: Busia District	-----Girls-----		Difference (s.e.)
	Program	Comparison	
Age in 2001	13.5	13.4	0.0 (0.1)
Father's education (years)	5.2	5.2	0.2 (0.5)
Mother's education (years)	4.6	4.6	0.1 (0.4)
Total children in household	7.0	6.5	0.5 (0.5)
Proportion ethnic Luhya	0.49	0.47	0.03 (0.05)
Latrine ownership	0.96	0.94	0.02 (0.01)
Iron roof ownership	0.77	0.77	0.00 (0.03)
Mosquito net ownership	0.33	0.33	0.00 (0.03)
Test Score 2000–Baseline sample (cohort 1 only)	-0.05	-0.12	0.07 (0.18)
Test Score 2000–Main sample (cohort 1 only)	0.07	0.03	0.04 (0.19)

Panel (A)



Why might incentives have an impact?

Theoretical perspectives

- Extrinsic motivation (exploiting immediate gratification)
- vs. Intrinsic motivation (“love of learning”)
- Great teacher effort (altruism, recognition)
- Parent encouragement / pressure on the girls
- Community mobilization to support the program

GSP empirical impacts (2001-2002)

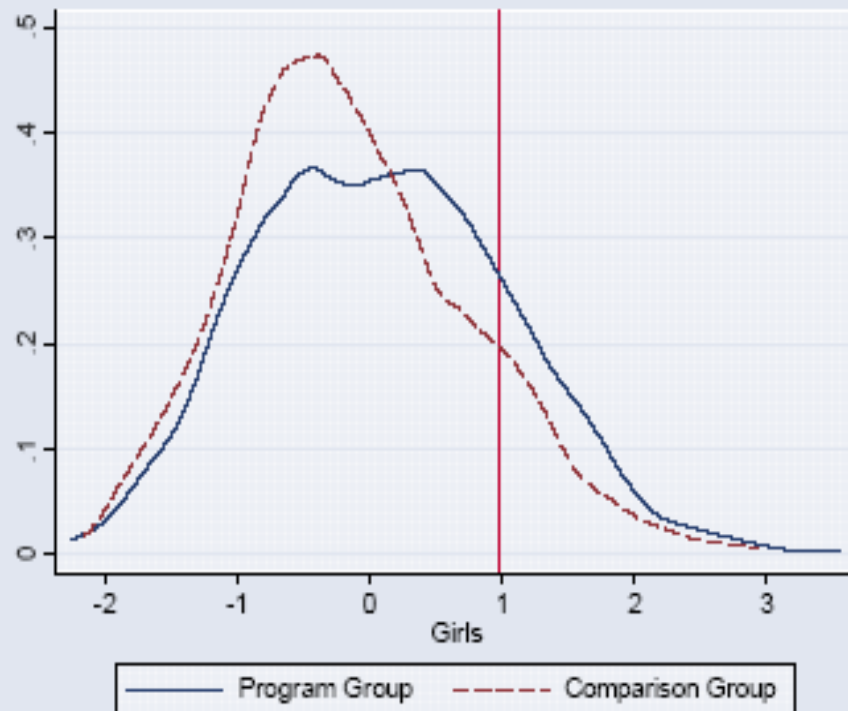
- Impacts are positive and quite large for cohort 1: 0.12-0.13 standard deviations on average (Table 4)
- There are positive effects for boys, too – even though they were not eligible for the prize: externalities
- Positive effects are concentrated in Busia district (gains of 0.2 s.d.), but are zero in Teso district – why?

Table 4: Program Impact on Test Scores
Longitudinal Sample, Cohort 1 Girls and Boys

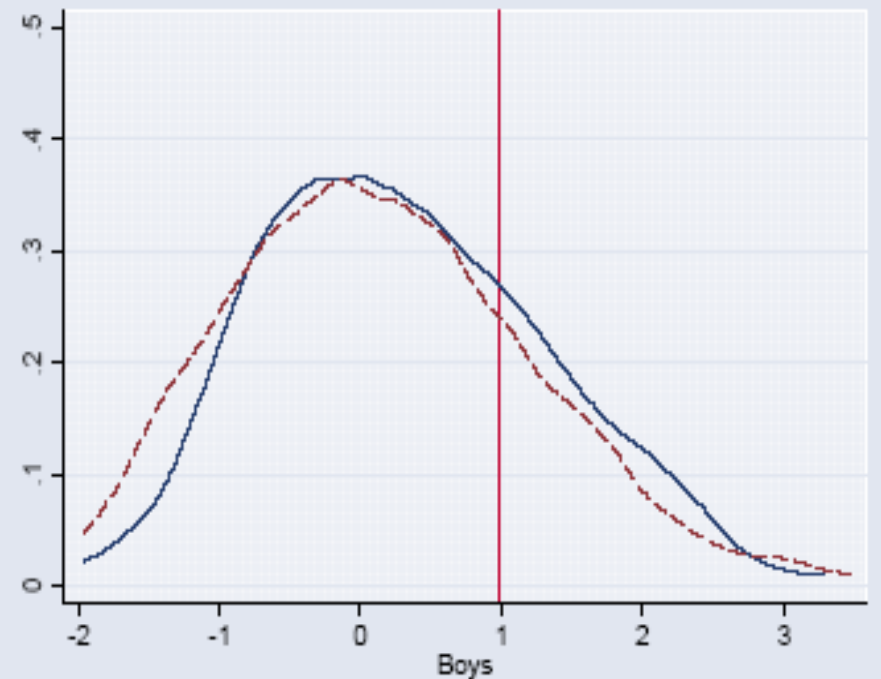
	Dependent variable:				
	Normalized test scores from 2001 and 2002				
	Busia and Teso districts		Busia district		Teso district
	(1)	(2)	(3)	(4)	(5)
Program school	0.12 (0.13)	0.13 ^{**} (0.06)	0.12 [*] (0.07)	0.19 ^{**} (0.08)	-0.02 (0.09)
Male * Program School			0.01 (0.05)	0.01 (0.05)	0.01 (0.09)
Male			0.16 ^{***} (0.04)	0.09 ^{**} (0.04)	0.28 ^{***} (0.07)
Individual test score, 2000		0.80 ^{***} (0.02)	0.79 ^{***} (0.02)	0.85 ^{***} (0.03)	0.69 ^{***} (0.02)
Sample Size	4294	4294	4294	2858	1436
R ²	0.00	0.61	0.61	0.67	0.53
Mean of dependent variable	0.13	0.13	0.13	0.13	0.12

Figure 6: Year 1 (2001) Test Score Distribution
Cohort 1 Busia Girls (Panel A) and Busia Boys (Panel B)
(Non-parametric kernel densities)

Panel (A)



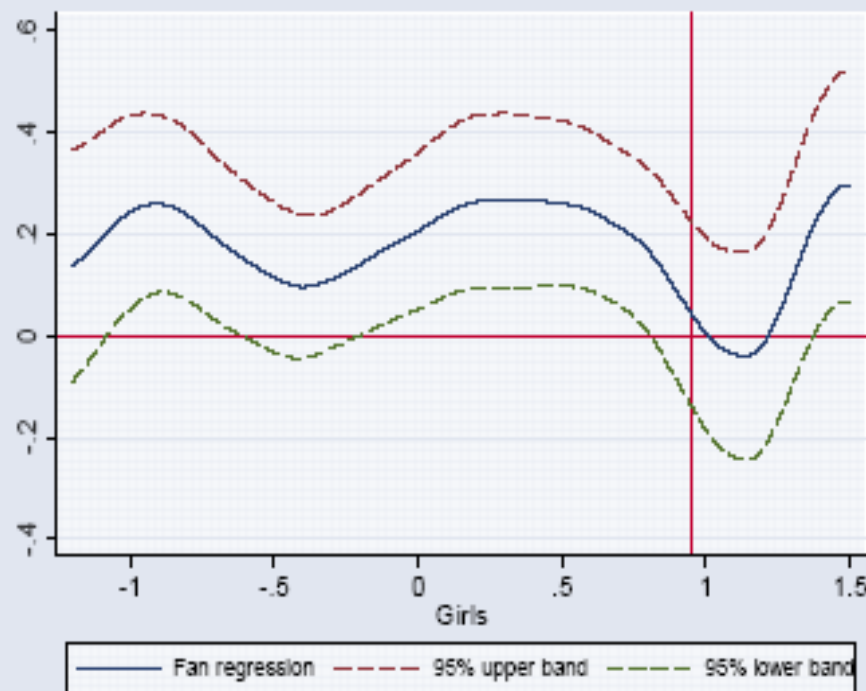
Panel (B)



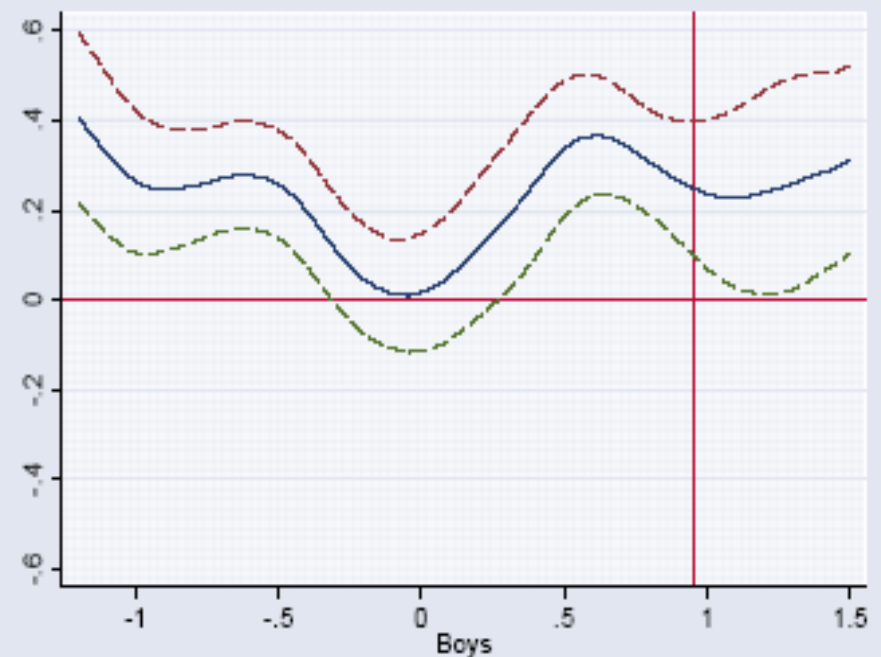
Vertical line represents the minimum winning score in 2001 in Busia.

Figure 7: Year 1 (2001) Test Score Impacts by Baseline (2000) Test Score Difference between Program Schools and Comparison Schools
Cohort 1 Busia Girls (Panel A) and Busia Boys (Panel B)
 (Non-parametric Fan locally weighted regression)

Panel (A)



Panel (B)



Vertical line represents the minimum winning score in 2001.

Difficulties in Teso district

- This NGO, and other NGOs, have long had trouble introducing new projects into Teso district
- The dominant ethnic groups are different in Busia district (Luhya) and Teso district (Teso)
- There was a **tragic lightning strike** incident in a Teso district primary school in April 2001 – seven students died (27 injured), and NGO project work became even more difficult afterwards. Five Teso district schools pulled out of the program

Figure 1: Map of Busia District and Teso District, Kenya, with location of Girls Scholarship Program Schools (legend below)

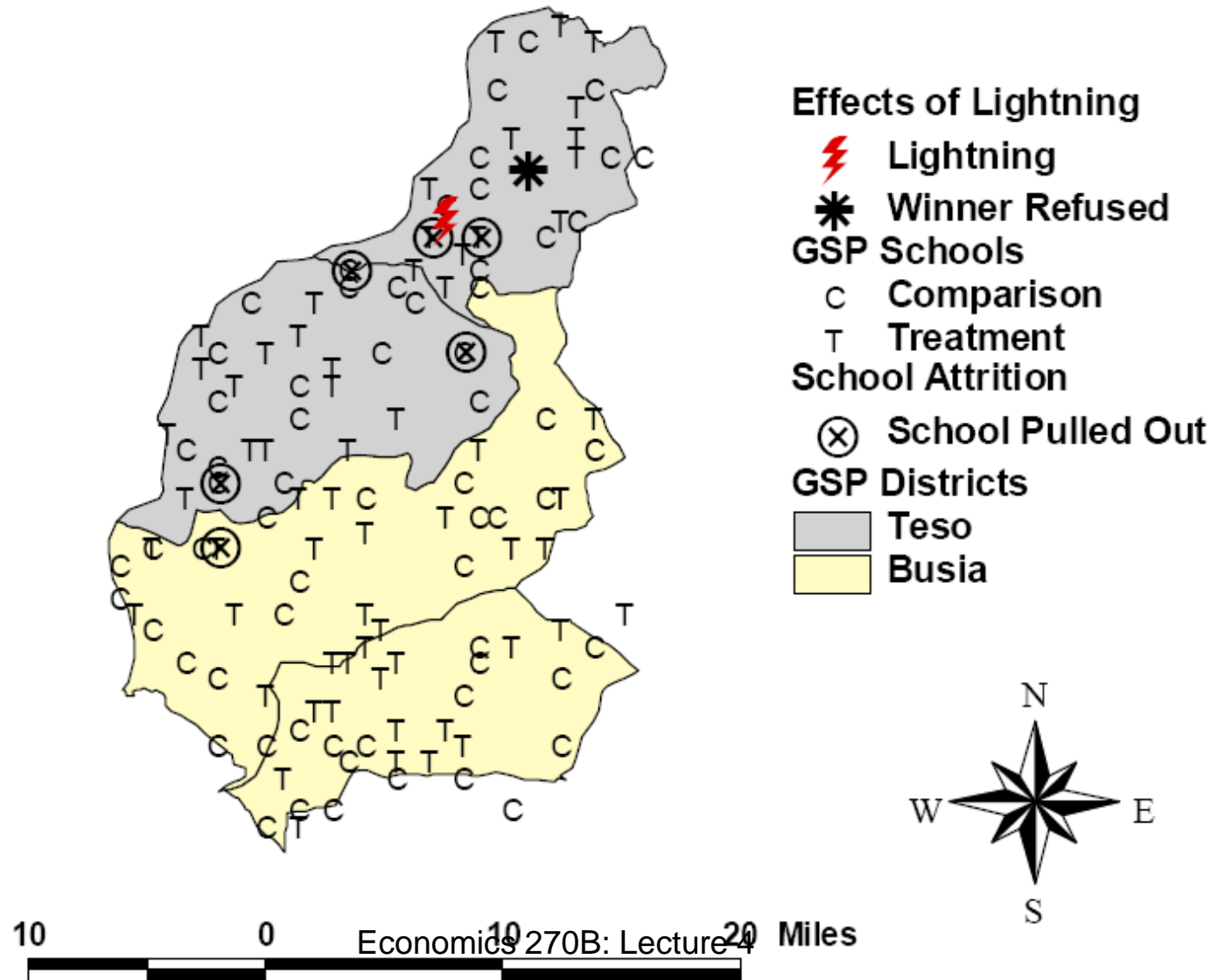
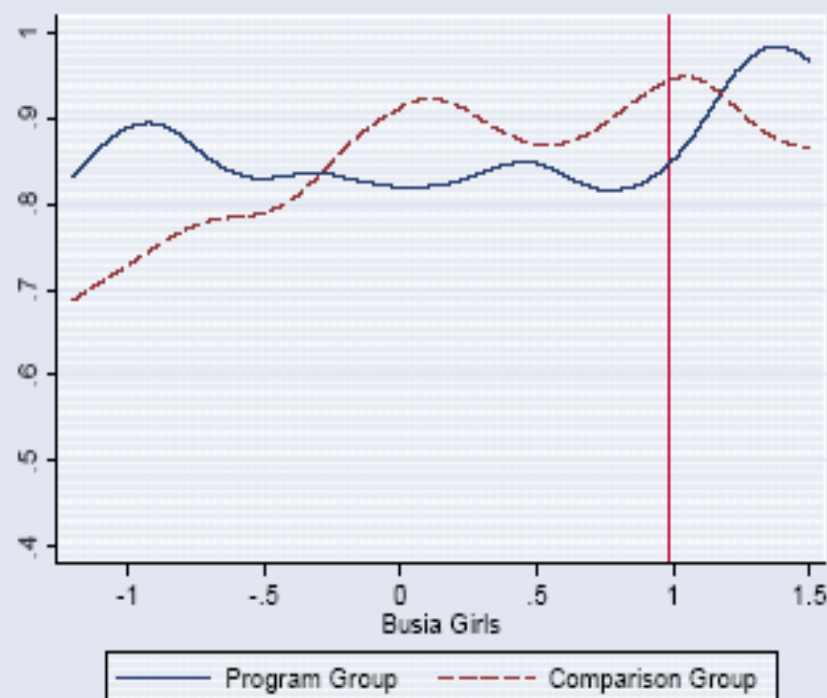
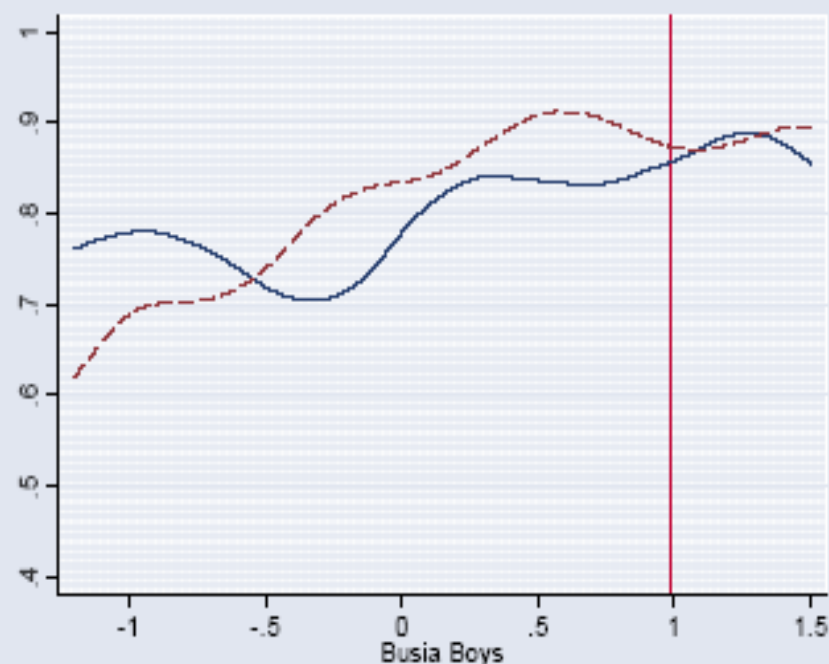


Figure 3: Proportion of Baseline Students in the 2001 Main sample by Baseline (2000) Test Score
 Cohort 1 Busia Girls (Panel A) and Busia Boys (Panel B)
 (Non-parametric Fan locally weighted regressions)

Panel (A)



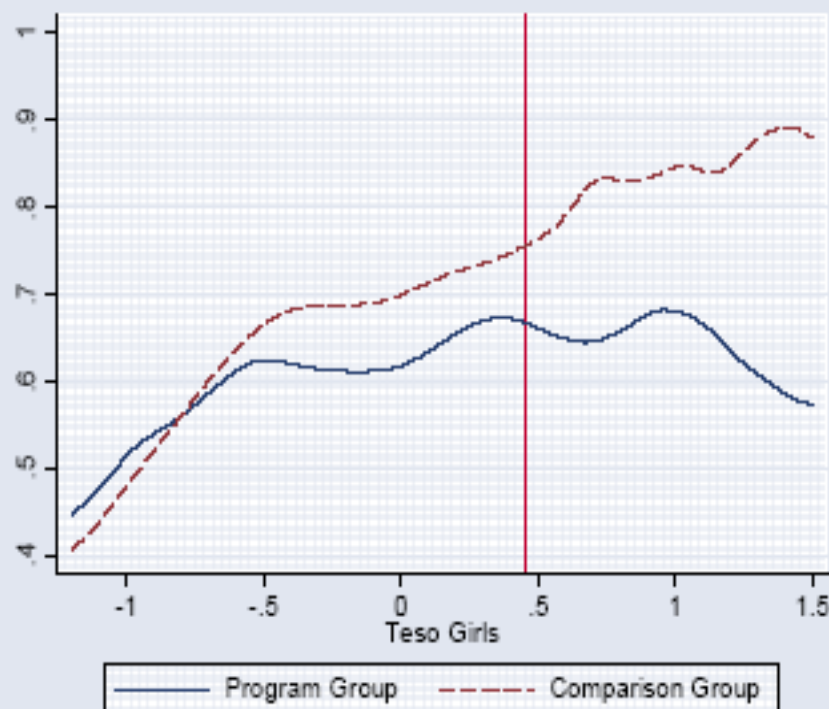
Panel (B)



Vertical line represents the minimum winning score in 2001.

Figure 4: Proportion of Baseline Students in the 2001 Main sample by Baseline (2000) Test Score
 Cohort 1 Teso Girls (Panel A) and Teso Boys (Panel B)
 (Non-parametric Fan locally weighted regressions)

Panel (A)



Panel (B)

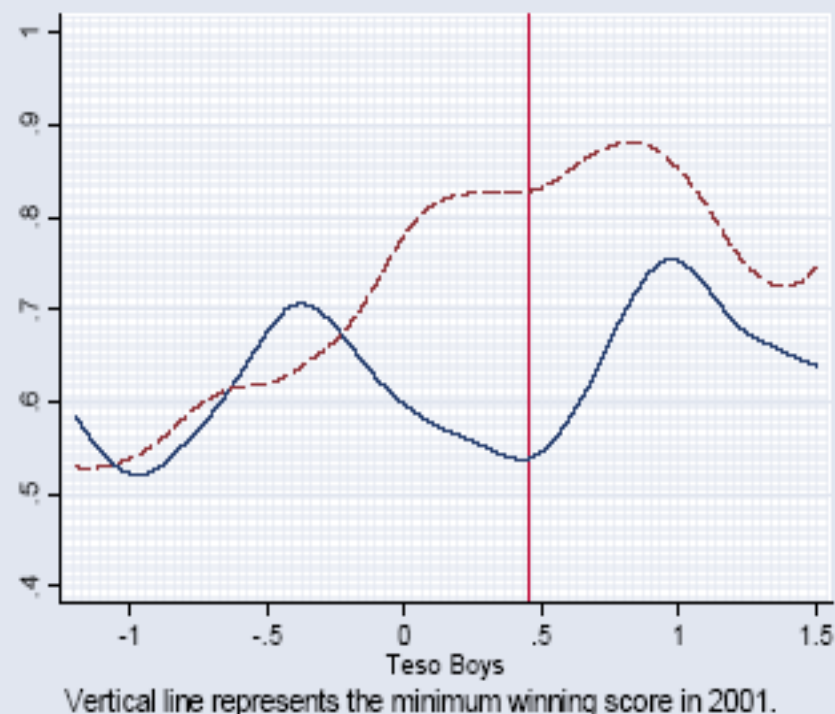


Table 5: Program Impact on Test Scores
Main sample, Cohorts 1 and 2 Girls and Boys

	<u>Dependent variable:</u>			
	Normalized test scores from 2001 and 2002			
	-----Girls-----		-----Boys-----	
	<u>Busia and Teso</u>	<u>Busia District</u>	<u>Busia and Teso</u>	<u>Busia District</u>
	(1)	(2)	(3)	(4)
Program year, Cohort 1 (2001)	0.18** (0.08)	0.28*** (0.10)	0.10 (0.07)	0.18** (0.09)
Program year, Cohort 2 (2002)	0.13* (0.07)	0.21** (0.10)	0.04 (0.10)	0.11 (0.13)
Post-competition year, Cohort 1 (2002)	0.12 (0.08)	0.25*** (0.09)	0.05 (0.07)	0.07 (0.09)
Mean school test score, 2000	0.75*** (0.05)	0.83*** (0.05)	0.78*** (0.06)	0.87*** (0.06)
Sample Size	4736	2917	5332	3206
R ²	0.29	0.36	0.26	0.32
Mean of dependent variable	-0.06	-0.03	0.21	0.21

Notes: Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. OLS regressions, Huber robust standard errors in parenthesis. Disturbance terms are allowed to be correlated across observations in the same school, but not across schools. Test scores were normalized such that comparison group test scores had mean zero and standard deviation one. Indicator variables are included in both specifications for Cohort 1 in 2001, Cohort 1 in 2002, and Cohort 2 in 2002 (coefficient estimates not shown). Main sample includes students who were registered in grade 6 (cohort 1) or grade 5 (cohort 2) in January 2001, in schools that did not pull out of the program, for whom we have mean school test score data in 2000, and who took the 2001 or 2002 test.

Evaluating critiques of merit scholarships

- No statistically significant changes in test score inequality in treatment schools
- Effort increased: student school participation increased by 5 percentage points in program schools (Table 7), for girls and boys in Busia district
- Teacher attendance increased 5 percentage points
- There are no significant changes in students' study habits, work at home, or attitudes toward education / stated intrinsic motivation (Table 6)

Dependent Variables:

Panel A: Attitudes towards education

Student prefers school to other activities (index) ^a

0.02 0.72
(0.01) (0.18)

Student thinks s/he is a "good student"

0.02 0.73
(0.04) (0.44)

Student thinks being a "good student" means "working hard"

-0.02 0.69
(0.03) (0.46)

Student thinks can be in top three in the class

0.00 0.33
(0.04) (0.47)

Panel B: Study/Work habits

★ Student went for extra coaching in last two days

-0.04 0.40
(0.04) (0.49)

Student used a textbook at home in last week

0.01 0.85
(0.03) (0.36)

Student did homework in last two days

0.03 0.78
(0.04) (0.41)

Teacher asked the student a question in class in last two days

0.03 0.81
(0.04) (0.39)

Amount of time did chores at home ^b

0.02 2.63
(0.05) (0.82)

Panel C: Educational Inputs

Number of textbooks at home

0.09 3.83
(0.19) (2.15)

Number of new books bought in last term

0.15 1.54
(0.14) (1.48)

What are the policy implications?

- Positive impacts:
 - Test scores improved considerably and for relatively low cost
 - GSP could promote empowerment of women and changes in social norms about girls' education
- Possible concerns / limitations:
 - Will the impacts last? In the long-run, will GSP really destroy the “love of learning” for these kids?
 - Does the lack of impacts in one of the two study districts indicate a relative lack of external validity, and caution in generalizing these findings to other settings?



Next week

- For next week's lecture, please focus on the Miguel and Kremer (2004) article.
- The third referee report is due in two weeks (March 9th), on the Morjaria article.