Policy Evaluation

Esther Duflo
September 5, 2012
IIE Anniversary Symposium

Introduction

- Policy evaluation was not very common in empirical development economics prior to the mid-1990s
- Program evaluation mainly absent from reading lists in the mid 1990s
- Classic empirical studies are not studies of policies:
  - Robert Townsend on village insurance in India
  - Chris Udry on insurance in Nigeria
- Even studies that have large policies looming in the background were not primarily interested in the impact of those policies:
  - Foster-Rosenzweig body of work on the green revolution
  - Munshi on contraception

Why was that the case?

- Empirical development economics was grounded in Schultz "Poor but efficient" paradigm
- Interest in estimation of "fundamental" parameters of the household models, e.g.:
  - Elasticity of food consumption with respect to outlay
  - Farm production functions
- Governments seemed to be a bit absent from households' life in this narrative
- To the extent policies were evaluated, it was to show that they are endogenously placed and have lower effect than anticipated.
  - Pitt Rosenzweig and Gibbons: Family planning in Indonesia.

Focus shifted due to a few key factors

- Body of theoretical work pointed poverty traps and role for government to bring change
  - Das Gupta and Ray, Stiglitz, Banerjee-Newman, etc.
- Macro work (Barro/Romer/Lucas) underscoring the importance of human capital in growth: what can governments do to foster human capital in practice?
- More interest in the role of governments per se (precursor to the recent boom in the study of political economy).
  - Inspiration from labor economics and public finance, where empirical methods for program evaluation had made great progress in the 1980s.
  - Availability of better and larger-scale data (censuses, DHS surveys, etc.) for developing countries

Policy Evaluation

- Following inspiration of labor, initial interest in retrospectively studying policies with "natural experiments", e.g.
  - Banerjee et al. Operation Barga (land reform)
  - Duflo (2001, 2004) INPRES School construction
  - Chattopadhyay and Duflo (2004) Quotas for women
- These papers have two objectives:

  (1) evaluate an actual (large, important) policy

  (2) shed some light on theory

RCTs

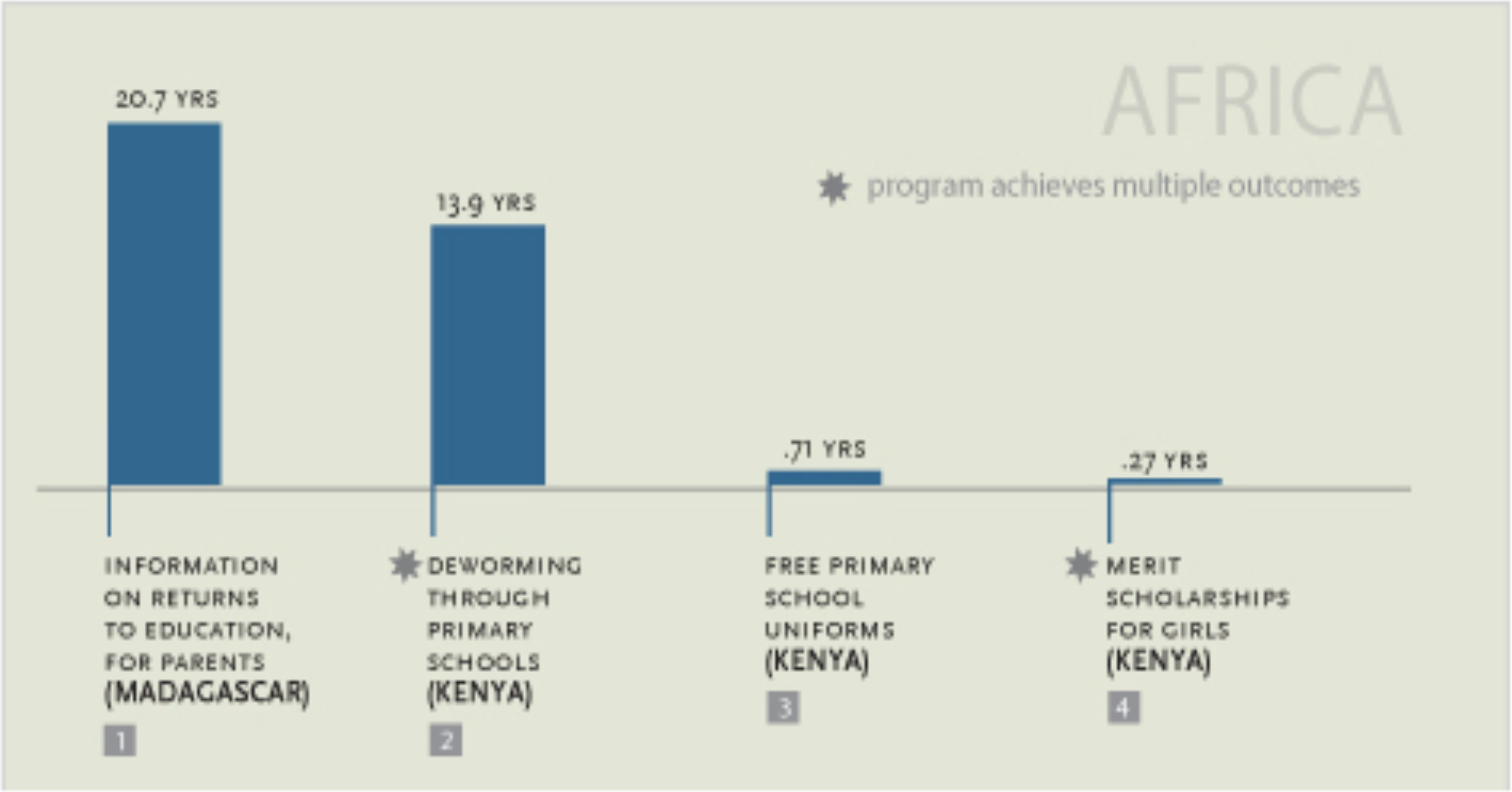
- At the same time of randomized controlled trials in development,
  - initiated by researchers: e.g. Kremer school projects in Kenya
  - And policy makers : e.g. PROGRESA in Mexico)
- Initial RCTs had relatively simple conceptual goals (evaluation of popular human capital policies such as textbooks, additional teachers, teacher incentives, etc.) while developing the practice of experiments in tricky contexts (managing clustering, attrition, imperfect compliance, etc.)

Key insights from these first evaluations

- Intuition or theory is NOT sufficient to predict what will and will not work.
  - Textbooks had no effect on average child's test score
  - Reducing class size had very small impacts on test scores
  - Deworming had large effect on attendance
  - Uniforms had fairly small effects
  - Costs/benefits of various policies is widely different from one program to the next: marginal rate of productivity of public funds is very far from being optimized!

Policy Evaluation: Education

COST-EFFECTIVENESS: ADDITIONAL YEARS OF STUDENT PARTICIPATION PER $100



Lessons

- Intuition does not provide an operational guide for what policy might do
- But the same is true of theory! Theory does not provide any guidance as to what magnitudes (or sign…) should be expected from these programs
- Ex-post findings can of course be rationalized and will seem obvious to someone….
  - if bednet price don't influence use neo-classical economists will find this obvious, behavioral economists may have found the opposite obvious).
- Clean estimation of particular components of policy is difficult in the real world: most policies come as packages and packages are not necessarily well-motivated

Policy Evaluation

- Best way to provide guidance for policy design is:

  (1) Evaluate past policies, particularly those that are precisely defined and have a single goal (i.e. school construction in Indonesia rather than CCT program that does many things at once)

  (2) Conduct RCTs that provide opportunity to experiment with single components or combinations

Why Evaluate Policy?

- Only way to know which policies work and which do not work, and why
- This raises two issues
  - Does that mean policy evaluation should be left to non-academics?
  - Do we care what works, if policy making is primarily dominated by politics, and hence knowing what works make no difference.

Should policy evaluation left to non-academics?

- Argued often: "leave it to the world bank".
- Why academics should be into this:
  - Not trivial: lots of methodological work needed to get it right (both non experimental methods and non experimental methods): the methods then percolate in policy world, and that's good.
  - Policies provide variation in data that provide researchers the challenge and opportunity to better understand the world

Challenge

- Consider all results on education quality from various program evaluations taken together:
  - Very low performance of students in school (low reading levels, etc.) while enrollment is increasing
  - Traditional inputs do not seem to help the average child but might help the brightest children
  - Parents and teachers reduce their effort in response to these programs (Das et al. for parents, Dupas et al. for teachers)
  - Remedial programs and targeted instruction (tracking) helps a lot
  - Computer-assisted learning can also help when students are put in front of very guided software
  - Private schools do not do tremendously better than govt schools (compared to, for example, remedial education)

- Which model of the education sector works to explain all of these results?

A model of the education sector

- Elite Bias in education, shared by
  - Teachers
  - Parents
  - Students
- Although *even in the current system* there appears to be benefits at all levels, perception is of higher return at higher levels.
- Teachers teach to the top and ignore the rest of the students, who are quickly losing

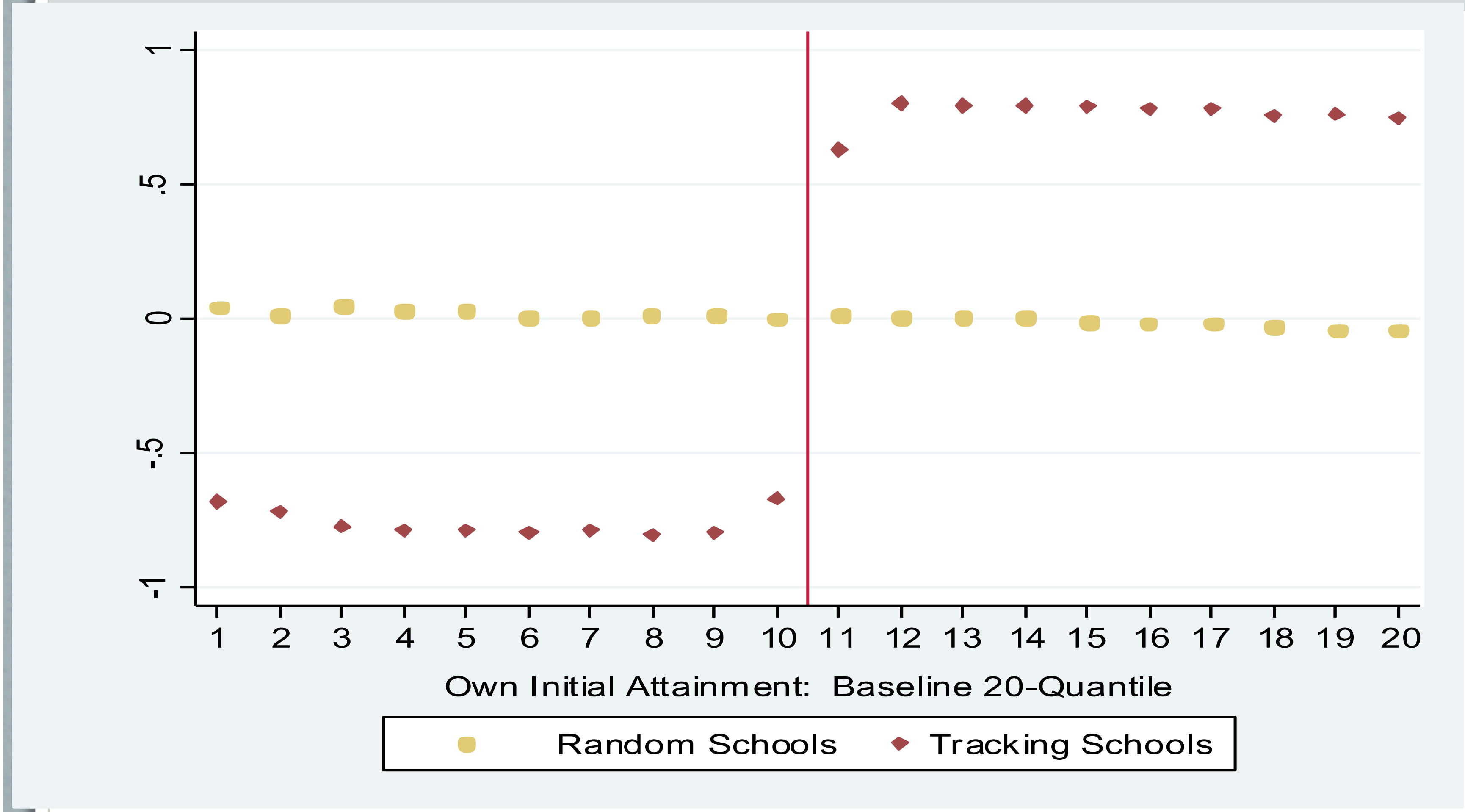
Opportunity

- Experiments can be designed to generate variation needed to test theories or hypotheses in richer ways than "naturally occurring" variation allows.

- For example, peer effects/social networks, diffusion effects are difficult to evaluate in "normal" conditions as people self-select into groups and neighborhoods, face common shocks. Random assignment to class groups avoids this problem.

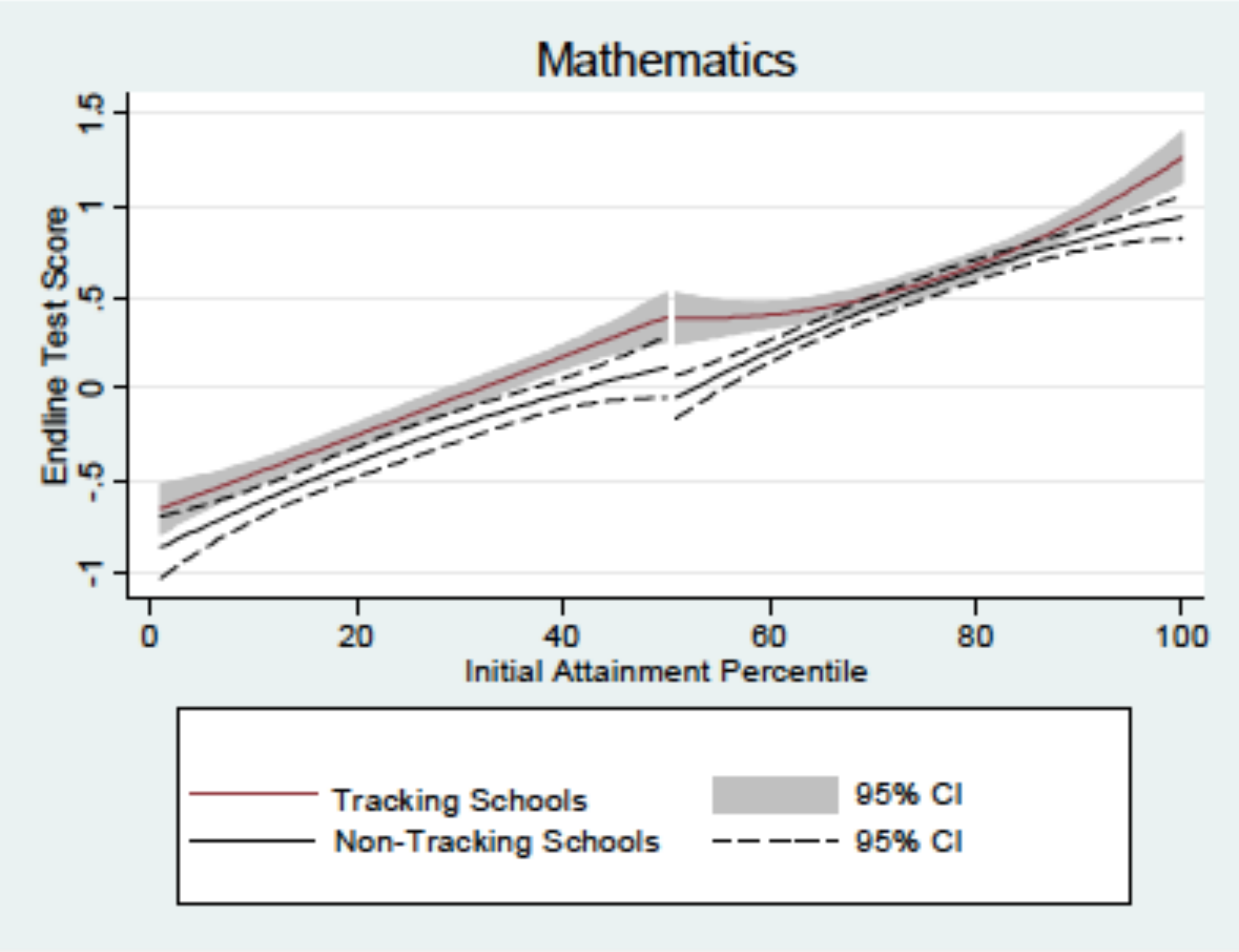
- Example: In Duflo, Dupas, Kremer (2011), students randomly assigned to a class group (control) and assigned according to prior grades in other group (treatment). Two opportunities to evaluate peer effecfs
  - Random assignment in the schools that were assigned to random school (identifies small variation)
  - Regression discontinuity at the median in tracking schools (identify combination of direct peer effect+change in teacher behavior).

Variation in Peer group quality



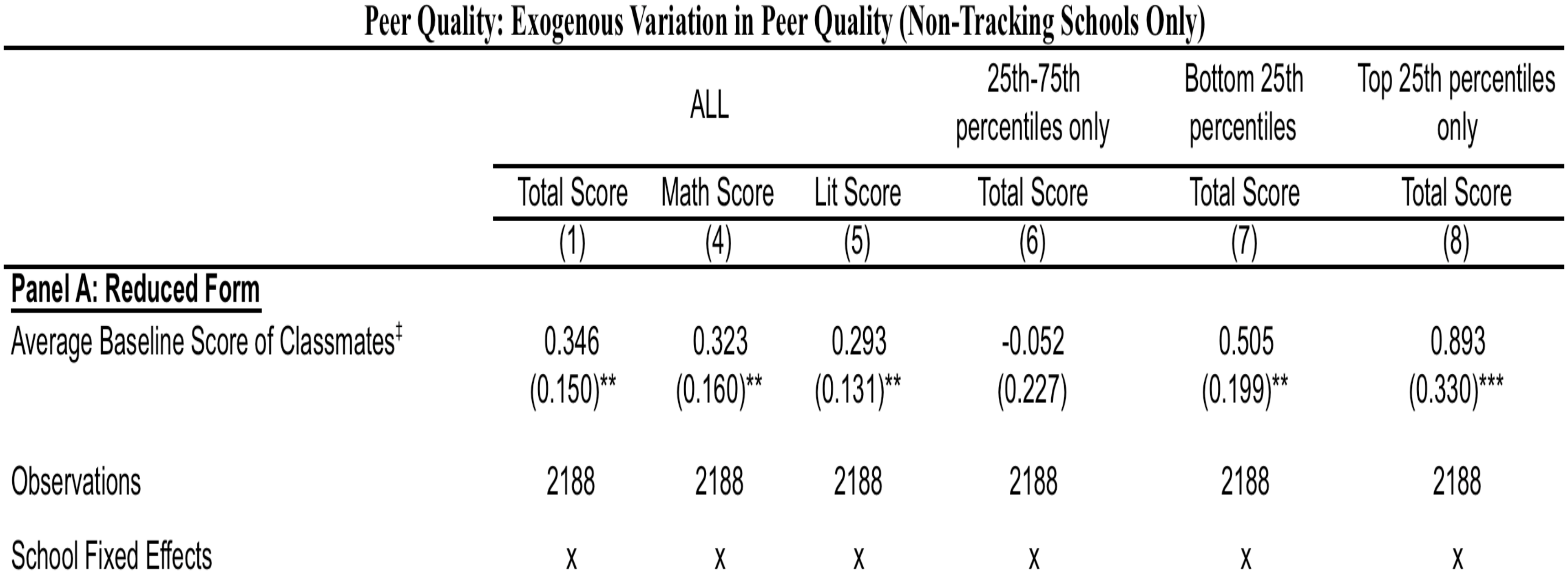
Peer Effects at the median

Local Polynomial Fits of Endline Scores by Initial Attainment



Source: Duflo, E., Dupas, P. and M. Kremer. "Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation"

Random variation in scores



Peer Quality: Exogenous Variation in Peer Quality (Non-Tracking Schools Only)

	ALL			25th-75th percentiles only	Bottom 25th percentiles	Top 25th percentiles only
	Total Score	Math Score	Lit Score	Total Score	Total Score	Total Score
	(1)	(4)	(5)	(6)	(7)	(8)
Panel A: Reduced Form						
Average Baseline Score of Classmates‡	0.346	0.323	0.293	-0.052	0.505	0.893
	(0.150)**	(0.160)**	(0.131)**	(0.227)	(0.199)**	(0.330)***
Observations	2188	2188	2188	2188	2188	2188
School Fixed Effects	x	x	x	x	x	x

Other Examples

- Karlan-Zinman:
  - Moral hazard and adverse selection in credit markets
  - Experimental design replicated in Ashraf et al, Dupas,
- Duflo, Kremer, Robinson, Schilbach (in progress):
  - Test epidemiological model of diffusion of information through networks
- Bertrand, Hanna, Mulainathan:
  - Corruption as "greasing the wheels" or distorting allocation

Should we bother?

- Alternative view of development that there is no point spending effort evaluating policies:
  - Development has to do with long run/"macro" things (institutions; religion)

- Understanding those is where the "high returns" are (or *may be*…) "And, the potential remedies for the fundamental problems holding back income growth might be much cheaper, in dollars, to implement than the set of programs in total promoted in Poor Economics" (Rosenzweig)

- May be there is not much we can do about these things:
  - Institutions are important, very slow moving, almost impossible to change by design (Acemoglu-Robinson)

"Big Answers for Big questions"

- The problem with the quest for the "big" determinants of growth has already been discussed elsewhere in the conference
  - We don't have good ways to evaluate these ideas (education in general, investment in general, trade in general, democracy in general, etc.)
  - And at the end of the day knowing, say, that "trade" is important will still be insufficient. That's an outcome. governments will need to know the impact of tariffs, barriers to international markets and how to get around them, etc.
    - Back to specific policy evaluations (RCT or not)
      - Topalova, Atkin, Atkin and Verhoogen, etc.
- Continuing Rosenzweig's quote "--the strategic removal of some red tape, de- or more efficient regulation are one-shot tweaks at the more macro level that might matter enormously for attracting "good" jobs and thus can have sustained future benefits" That's screaming for some for policy evaluation! (Besley-Burgess)

Institutions

- Institutions set some constraints on what policies will be pursued, and will be successful, but not a straightjacket
  - There is scope for good policies even in bad environments (e.g. Indonesia), and even for marginal improvements in institutional environments through policies (e.g. Olken's corruption experiment in Indonesia)
  - There is (plenty of) scope for bad policies in generally well functioning environments, precisely because
    - We know so little (ex ante) about what may work
    - Without explicit, rigorous, *evaluation it is not possible to know what has worked.* The democratic "marketplace" cannot play its role without evaluation. Free market for ideas is not a substitute for evaluation!

Information and the Vitality of Democracy

- Voters are sensitive to politicians' performance if they know about it:
    - Finnan and Ferraz (audits in Brazil)
    - Banerjee et al. (report cards for politicians in Delhi)
- Voters are able to have informed discussions about common decisions, and they like it.
    - Wanchekon (deliberative democracy, Benin)
    - Olken (town meeting versus election, Indonesia)

From Policy Evaluation to Policy Design

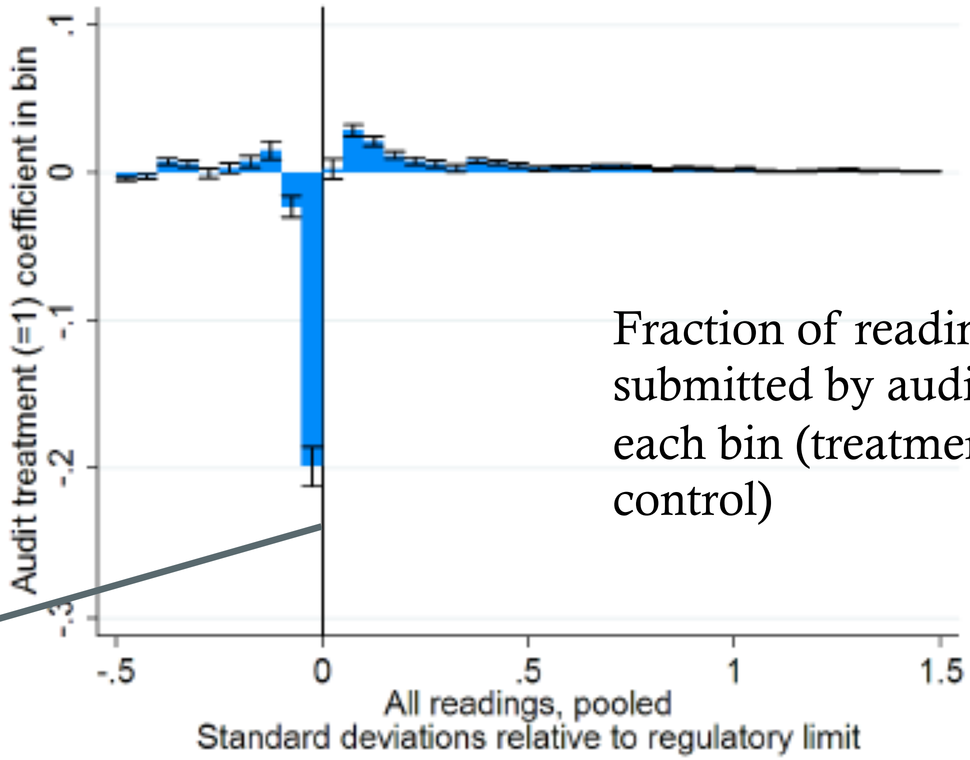
- Previously, trade-offs between "large-scale", black box, policy evaluations (e.g. PROGRESA) and more "boutique" experiments carefully tailored by researchers to test a particular theory.

- Trade-off is often no longer necessary: researchers can now work with governments to evaluate meaningful policies on large scale and shed light on interesting economic questions.

Third party auditors in India

- Work with govt of Gujarat to reform existing environmental scheme.
- Status quo: firms choose and pay auditors, must obtain one audit every year.
  - Firm pick auditors who will give good score to perform audit, and no one pays attention to the audits
- Very familiar problem: Credit rating, financial audits, etc..
- Typical proposed reform is to assign auditors to firms, with mandatory rotation or better monitoring of monitors
- Gujarat Experiment:
  - Auditors paid from central pool of funds, assigned to firms, and monitored by independent back-checks.

Auditors much less likely to report firms to be compliant



Fraction of reading submitted by auditor in each bin (treatment – control)

Regulatory Threshold

Source: Duflo, E., Greenstone, M., Pande, R., and N. Ryan. "Truth-Telling by Third-Party Auditors: Evidence from a Randomized Field Experiment in India."

Third Party auditors

- Found
  - auditors report more truthfully
  - firms pollute less.
- Two consequences:
  (1) administration is working to modify system (in Gujrat and elsewhere in India) and

  (2) insight on market for 3rd-party auditing

Targeting the Poor in Indonesia

- Alatas, Banerjee, Hanna, Olken, Tobias (2012)
  - What works best: administrative targeting or allowing the community to decide?
  - Communities pick the very poorest people better, for the one near the poverty line, targeting is worst, but they chose people who people tend to think are poor (may be not based on consumption)
  - They are much happier
- Response to the research
  - Process for updating the list of beneficiaries for the free rice program takes into account the result of the research. It is now possible to kick out someone the community considers to be "rich" and to replace them with someone else who is poor.
  - Use variation to estimate difference in beneficiaries and structural model of how information on people's poverty status diffuses in the network

Other Examples

- Series of CCT studies done with government that analyze various features (conditionality, gender of the recipient, levels)
- Dal Bo, Finnan and Rossi: randomized the wages of government employees in Mexico.
- Kremer and Mularidharan (in progress): randomization at individual-level and market-level of voucher to measure direct effect of private schools and potential effects on remaining children
  - Allows researchers to work on scale large enough to detect “market-level” impacts, blurring the distinction micro/ macro.

Discussion

- Focus remains on "micro" policies as opposed to country-wide system reforms that cannot be easily evaluated

- Some effort to bridge the gap is happening:
  - Predict/calibrate macro growth from micro parameter (Townsend)
  - Adapt natural experiment or RCT methodology to large scale context (e.g. Kremer-Muralidharan on voucher for private schools).

- No distinction between "pro-poor" and "pro-growth"
  - "Pro-growth" policies can be evaluated; not necessarily for impact on aggregate growth but for impact on channels that are supposedly at work in creating this growth.
  - Examples: industrial policy (support for some types of firms), trade policy (impact of exporting firms on employees, impact of policies to increase trade on trade), agricultural policy ("green revolution" steps such as technology development, adoption, etc.)