

Pre-Specification Across Research Projects

Thad Dunning
UC Berkeley, Political Science

Summer Institute
June 2014



Is Community Monitoring Effective?



Source: <http://cec.vcn.bc.ca/cmp/modules/mon-wht.htm>

June 2014

BITSS Summer Institute

Is Community Monitoring Effective?

This paper presents a randomized field experiment on community-based monitoring of public primary health care providers in Uganda. Through two rounds of village meetings, localized nongovernmental organizations encouraged communities to be more involved with the state of health service provision and strengthened their capacity to hold their local health providers to account for performance. A year after the intervention, treatment communities are more involved in monitoring the provider, and the health workers appear to exert higher effort to serve the community. We document large increases in utilization and improved health outcomes—reduced child mortality and increased child weight—that compare favorably to some of the more successful community-based intervention trials reported in the medical literature.

Source: Martina Bjorkman and Jakob Svensson, 2009, “Power to the People: Evidence from a Randomized Field Experiment on Community_Based Monitoring in Uganda.” *Quarterly Journal of Economics* 124 (2): 735-69.

Is Community Monitoring Effective?

This paper presents a randomized field experiment on reducing corruption in over 600 Indonesian village road projects. I find that increasing government audits from 4 percent of projects to 100 percent reduced missing expenditures, as measured by discrepancies between official project costs and an independent engineers' estimate of costs, by eight percentage points. By contrast, increasing grassroots participation in monitoring had little average impact, reducing missing expenditures only in situations with limited free-rider problems and limited elite capture. Overall, the results suggest that traditional top-down monitoring can play an important role in reducing corruption, even in a highly corrupt environment.

Source: Benjamin Olken, 2007, "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* : 115 (2): 200-49.

Is Community Monitoring Effective?

We study a randomized educational intervention in 550 households in 26 matched villages in two Kenyan districts. The intervention provided parents with information about their children's performance on literacy and numeracy tests, and materials about how to become more involved in improving their children's learning. ~~We find the provision of such information had no discernible impact on~~ either private or collective action. In discussing these findings, we articulate a causal chain linking information provision to changes in citizens' behavior, and assess the present intervention at each step. Future research on information provision should pay greater attention to this causal chain.

Source: Evan S. Lieberman, Daniel N. Posner, and Lily L. Tsai, 2013, "Does Information Lead To More Active Citizenship? Evidence from an Education Intervention in Rural Kenya." MIT Political Science, Working Paper No. 2013-2.

Why are estimated effects of community monitoring so different?

- One possibility: “chance” variation?
 - But, publication and reporting biases...
 - We likely don’t see the true distribution of estimated effects
- Some other possible answers:
 - The interventions are different
 - The outcomes are different
 - “It depends”

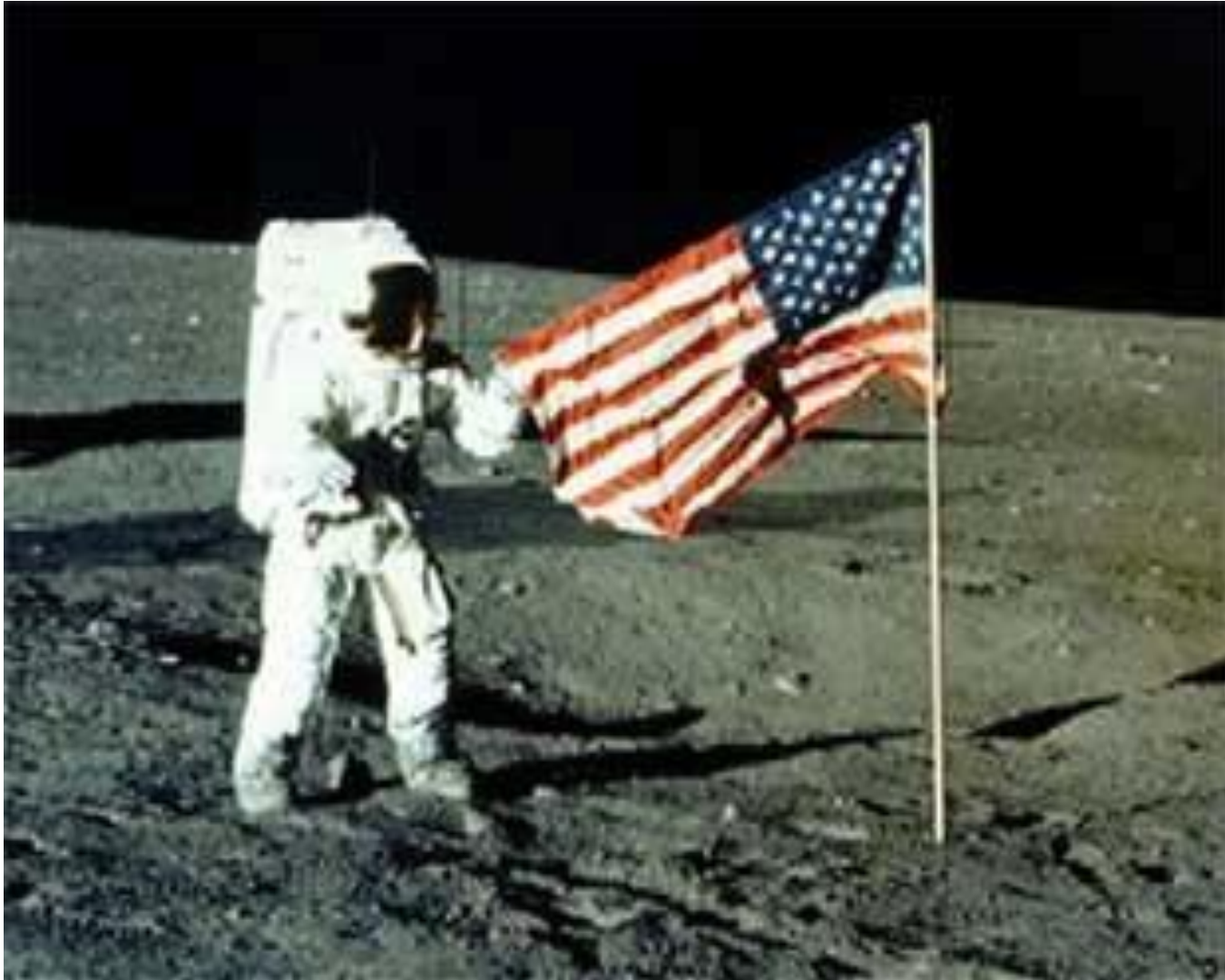
Some challenges for experimental social science

- The “credibility revolution” has increased the reliability of claims about causal effects.
- Yet several challenges remain, including difficulties of
 1. Achieving cumulative knowledge;
 2. Ensuring standards of analysis and reporting equal those of design; and
 3. Creating usable evidence for policy.

Overview

- Three challenges in more detail
- Pre-specification across research projects: a pilot initiative
- Strengths and limitations of this initiative
- Implications of collaboration for researchers

1. Challenges to Cumulation



1. Challenges to Cumulation

- Researchers often work independently, developing and addressing research questions that interest them.
 - Incentives to replicate previous research are often weak: too much “trust” and not enough “verify”
 - Broad conclusions are sometimes drawn from a single pioneering study.
 - Rewarding “planting the flag” is a source of publication bias—if follow-up null effects are harder to publish.
- Uncoordinated innovation, while laudable, can also hamper assessment of external validity
 - We’d like to understand what works in what contexts, and for what reasons.

2. Reporting Standards



2. Reporting Standards

- Without strong reporting standards, we risk undermining inferential gains from stronger designs
 - Estimates of effects in individual studies are more credible—but are bodies of literatures as a whole reliable?
- **Publication bias** – journals publish research that shows statistically significant effects
 - Distribution of published effects does not represent the distribution of true effects
 - But null effects are not null findings!
- **Multiple comparisons**—but “single reporting”
 - Nominal p-values don’t represent the true probabilities

Evidence of publication bias (Gerber and Malhotra 2008)

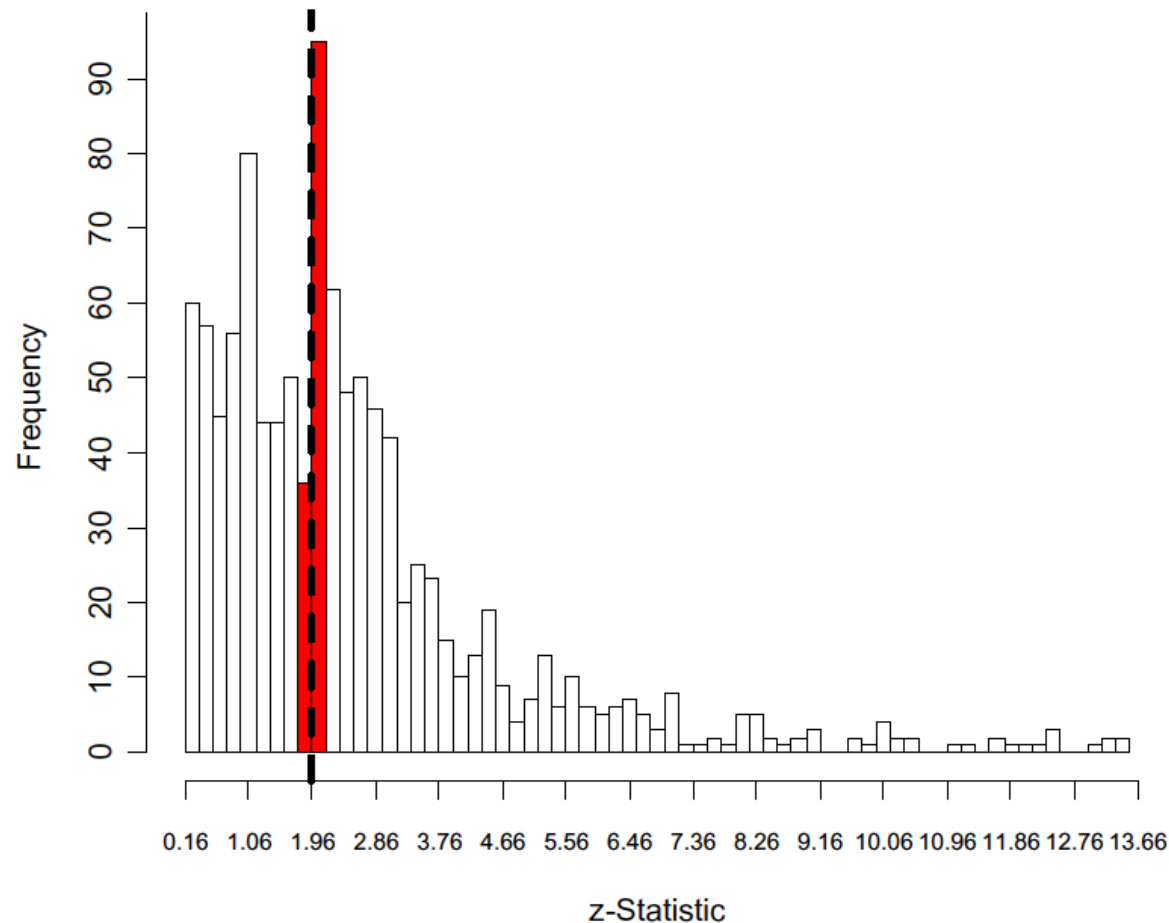


Figure 1(a). Histogram of z -statistics, *APSR* & *AJP* (Two-Tailed). Width of bars (0.20) approximately represents 10% caliper. Dotted line represents critical z -statistic (1.96) associated with $p = 0.05$ significance level for one-tailed tests.

Evidence of publication bias (Gerber, Green, Nickerson 2001)

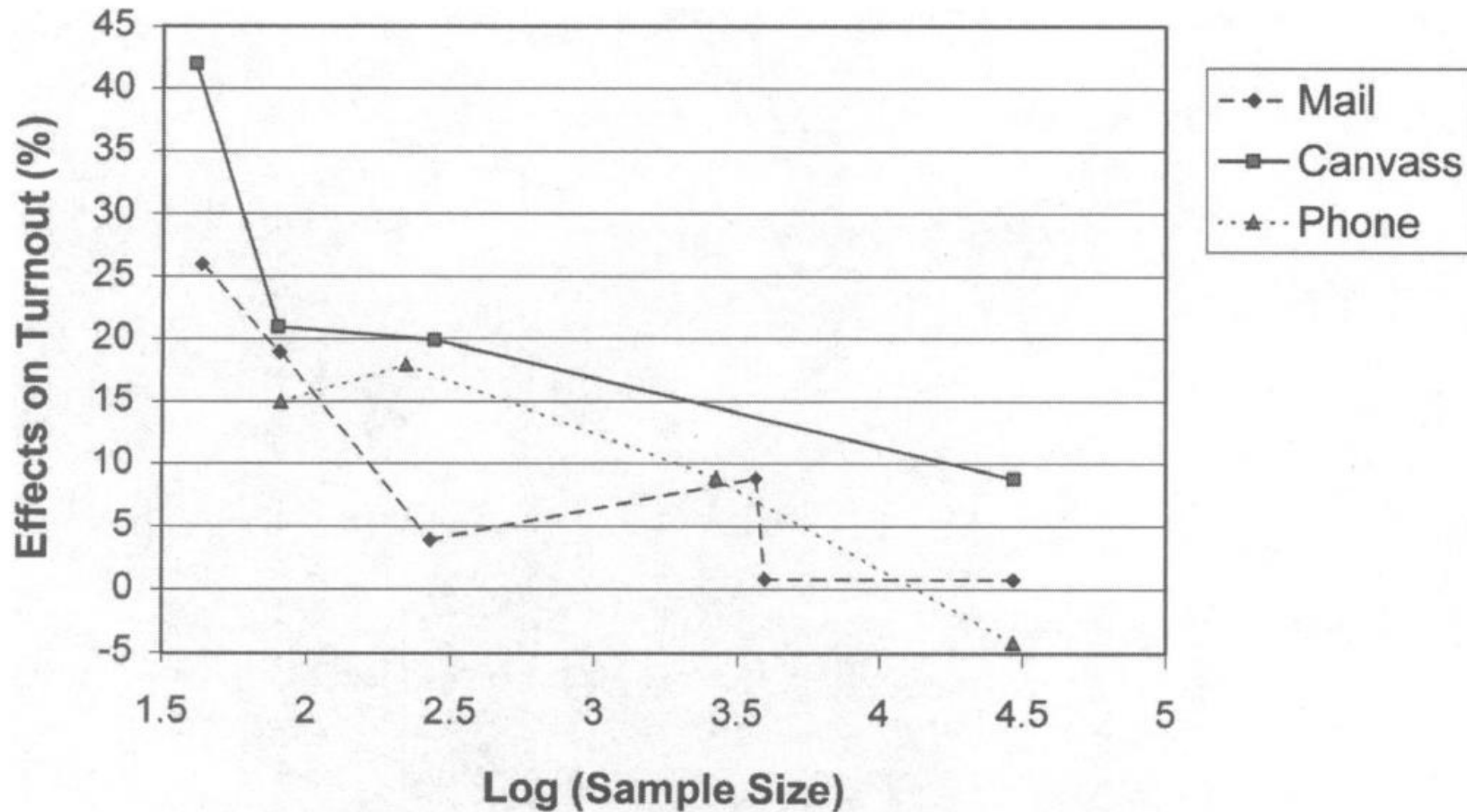


Fig. 1 Relationship between sample size and effect size.

Some potential fixes

- **Study registration**
 - Allows description of universe of studies
 - But also leaves substantial researcher degrees of freedom
- **Pre-analysis plans**
 - Limits data mining and permits meaningful adjustment for multiple statistical comparisons
 - But does not necessarily limit publication bias
- **Results-blind review**
 - Allows evaluation based on the quality of the research question and strength of the design – not the statistical significance of estimated effects
 - A potentially powerful tool for limiting publication bias (but not practiced yet); some potential drawbacks but not insurmountable

But what about synthesis?

- For pooling the results of multiple studies (e.g., meta analysis), there remains significant discretion and uncertainty.
 - What is the universe of studies?
 - Are interventions and outcome measures comparable?
 - Are we estimating the same parameter with different subjects in each study—or different parameters?
- Meta-analysis presumes conditions that are often unlikely to be met in practice
- Difficulties for synthesis can also be traced to uncoordinated innovation and challenges for cumulation

3. Creating Usable Knowledge



3. Creating Usable Knowledge

- Uncertainties regarding external validity can make it difficult to import knowledge from one context into another and can provide an avenue for discounting unwelcome findings.
 - Effects may be heterogeneous across contexts or countries—yet features of contexts are not manipulated or even manipulable.
- Despite difficulties, it seems critical to explore whether channels that link interventions to outcomes are operative in different contexts
- A framework for specifying and validating ex-ante predictions about heterogeneous effects may be helpful.

A Pilot Model: EGAP Regranting Initiative

- The Experiments in Governance and Politics (EGAP) group is running a three-year, \$1.8M regranting window, housed at Berkeley's Center on the Politics of Development (CPD).
- Objective: to pilot a model for experimental research that may address these key challenges
- A central difficulty:
 - How to foster greater integration of research projects, while getting researcher incentives right?
- Changing the funding and publication model may help

Pre-specifying a focus

- Two-stage process:
 1. Expression of Interest (with several possible themes identified)– used to identify promising clusters/thematic areas for proposals.
 2. Request for Proposals – proposals due June 16 (see e-gap.org or cpd.berkeley.edu)
- Criteria for selecting thematic focus in stage 1:
 - Previous body of research exists
 - Candidate interventions that are tested, scalable, simple, portable, punctual, ethical (!)
 - Capacity for analysis of downstream and heterogeneous
 - Some feasibility concerns (e.g., three-year grant window)
 - Funder priorities (to some extent)

Thematic Focus: Citizen Engagement And Political Accountability



Theoretical Focus 1: Informational Interventions

- Why do voters select underperforming politicians?
 - A growing body of research focuses on effect of *informational interventions* on voter behavior.
- Results to date are mixed – but not easy to understand sources of heterogeneity (due inter alia to variations in treatments and outcomes)
- Tractable area for three-year grant window, e.g. due to focus on elections.
- Also largest area for Expressions of Interest.
 - Quite interesting convergence across unrelated proposals.
 - Outside of this initiative, researchers might conceivably worry about being “scooped”
 - Participation in a joint project with integrated publication may help ease those concerns, to some extent.

Theoretical Focus 2: Information vs. Alternatives

- We want to build in replication and cumulation—but also make the initiative appealing to researchers
 - Also, some discomfort with sole focus on informational effects.
- The RFP thus specifies two treatment arms:
 - An informational arm that is consistent across all studies.
 - An alternative intervention that could be informational (with variation in treatment), or could be something else.
- This structure promotes replication and comparability—through the first treatment arm—while preserving room for innovation through the second arm.
 - We hope this helps to get researcher incentives right.

Seven pillars to the approach

1. Predefined themes.
2. Coordination and competition.
3. Comparable interventions and outcome measures.
4. Preregistration.
5. Third-party analysis.
6. Formal synthesis based on ex-ante planning.
7. Integrated publication -- and perhaps results-blind review.

Next steps on the regranting initiative

- There are a number of difficulties:
 - Capacity to generate integrated projects is untested; failure rate of individual studies may be high.
 - Small numbers of projects funded in relatively small amounts; so scope for meta analysis is still limited.
- But we received a large number of Expressions of Interest (61 in all), suggesting several interesting clusters.
 - We hope this can lay the groundwork for future funding rounds, as we move beyond this pilot initiative.
- Next steps after awards – workshop designs and harmonize interventions and outcomes
 - Collaborative theory (e.g. of heterogeneous effects)
 - Joint pre-analysis plan (for “study of studies”)

Strengths of Shared Research

- Our hope is that this model can
 - ***Foster cumulation.*** Group proposals (or grouped individual proposals) will address similar questions, introduce variation in treatments in systematic fashion, and coordinate outcome measures.
 - ***Improve synthetic analysis.*** Pre-registration of groups of studies defines the universe of comparisons.
 - ***Help illuminate what works where and why.*** Case selection, and theory about why and where we should see heterogeneous effects, is a critical part of proposals; we want to validate these predictions and assess when key channels are operative.
- Getting researcher incentives right seems critical.
 - Integrated publication may help.

Limitations of Synthetic Analysis

- With pre-specification across projects, estimating “SATE” for study group is straightforward
 - Like a large experiment with assignment blocked by country or research site
 - True without pre-specification—**but** with joint pre-planning, much greater harmonization of interventions and outcomes.
 - This is critical for meaningful synthesis
- But no panacea for synthesis...
 - Is the study group a “sample”? What is the population?
 - In particular, what is the population estimand we’d like to estimate?
 - E.g., average vs. heterogeneous effects

Understanding what works where, and why

- Essentially, a question about mechanisms
 - But analysis of mechanisms/mediators always very tricky
- Variations in treatment provide some opportunities
 - Explore what component of treatment is effective
- EGAP regranting initiative leaves scope for variation in informational interventions
- Can variation in treatments across studies illuminate mechanisms?
 - Perhaps, through design choices and a mix of methods

Cumulation through Variations in Treatment

The recent wave of randomized trials in development economics has provoked criticisms regarding external validity. We investigate two concerns—heterogeneity across beneficiaries and implementers—in a randomized trial of contract teachers in Kenyan schools. The intervention, previously shown to raise test scores in NGO-led trials in Western Kenya and parts of India, was replicated across all Kenyan provinces by an NGO and the government. Strong effects of short-term contracts produced in controlled experimental settings are lost in weak public institutions: NGO implementation produces a positive effect on test scores across diverse contexts, while government implementation yields zero effect. The data suggests that the stark contrast in success between the government and NGO arm can be traced back to implementation constraints and political economy forces put in motion as the program went to scale.

Source: Tessa Bold, Mwangi Kimenyi, Germano Mwabu, Alice Ng'ang'a, and Justin Sandefur, 2013, "Scaling Up What Works: Experimental Evidence on External Validity in Kenyan Education." Center for Global Development, Working Paper 321.

Implications for researchers

- Especially for researchers early in their careers: is there any conflict between this scientific model and professional advancement?
 - Innovation is critical for research
 - It is also highly professionally rewarded
- Is the model scalable?
 - It might be attractive because it is somewhat novel!
- This model combines replication and innovation
 - E.g. experimental designs with variations in treatment
 - Replication arms and “innovation” arms
- We hope this helps to reconcile professional and scientific rewards