



THE WORLD BANK



Measuring Results and Impact Evaluation: From Promises into Evidence

**Dean Karlan
Yale University
Innovations for Poverty Action**

Human Development
Network

East Asia and the Pacific
Region

Spanish Impact Evaluation
Fund

How do we turn this teacher...



...into this teacher?



05.01.2006 12:00

The Skeptics of Aid

“After \$2.3 trillion over 5 decades, why are the desperate needs of the world's poor still so tragically unmet? Isn't it finally time for an end to the impunity of foreign aid?”

- **Bill Easterly**

Skeptics believe that “the system” is so broken that positive change is impossible. How do we address this concern?



Maximizing impact under constraints

- How much would you spend to save a person drowning in a lake?
 - \$50? \$100? \$5000?
 - Are they worth saving if it costs \$1,000,000?
 - What if it costs \$1,000,000 to save a drowning person, and \$500,000 to save someone trapped in a burning building
- In the context of development, rigorous evaluations are the key to resolving such conundrums



Answer Three Questions:

1. Why is evaluation valuable?
2. What makes a good impact evaluation?
3. How to implement an impact evaluation?

Why Evaluate?

- ❑ Overall theme: Think forward, not backwards
- ❑ Need evidence on what works
 - Development resources are limited.
 - Without evidence on what works, there is no way to determine optimal allocation
 - Poor allocation of scarce resources has a real cost, often borne in the form of lost lives, futures, and income
- ❑ Improve program/policy implementation
 - Evaluation can help improve existing programs
 - Operations: efficiency & targeting
- ❑ Information key to sustainability
 - Budget negotiations
 - Informing beliefs and the press
 - Results agenda & Aid effectiveness

Impact Evaluation Answers

- ❑ What was the effect of the program on relevant outcomes (test scores, health status, income, etc)?
- ❑ How much better off are the beneficiaries because of the program/policy?
- ❑ Is the program cost-effective?
- ❑ Which programs deliver the most “bang for the buck?”
- ❑ Traditional M&E cannot answer these

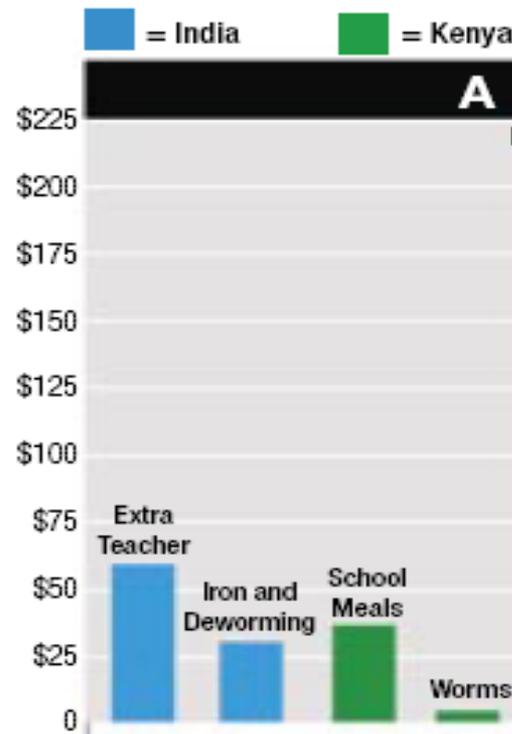
A basis for optimal allocation

- Suppose I have a million dollar endowment to spend on increasing primary school attendance. I can allocate it for one of these four projects...
 - Free meals for students in Kenya
 - Free deworming treatment for students in Kenya
 - Hiring of extra teachers in India
 - Free deworming treatment and iron fortification for students in India
- How do I know which of these investments will generate the maximum returns in the form of increased attendance?



Cost-Benefit Analysis

- Suppose you have the following results from programs evaluated in India and Kenya



Cost per additional year of schooling induced

From this chart, it is evident that the money will have the greatest impact on attendance if spent on deworming treatment for students in Kenya



How to allocate limited resources?

- Cost-Benefit analysis
 - allows comparison of choices
 - indicates highest return investment
 - Enables optimal resource allocation
- Cost-Benefit analysis needs rigorous evaluations.
 - Traditional forms of monitoring and evaluation have difficulty generating reliable figures such as the ones just displayed.
 - There is therefore a clear need for rigorous impact evaluation of programs

Impact evaluation helps us:

- ❑ Decide whether pilot-interventions/programs are worth scaling up...
- ❑ Decide whether programs should be discontinued...
- ❑ Modify existing programs to maximize their impact...
- ❑ Perform rigorous cost-benefit analysis

Answer Three Questions:

1. Why is evaluation valuable?
2. What makes a good impact evaluation?
3. How to implement an impact evaluation?

How to assess impact

- e.g. Do free lunches improve student test scores (learning)?
- We must compare:
 - How a student scores when he receives free lunches
 - How a student would have scored without the free lunches (counterfactual)
- How do we create a counterfactual?
 - We must give some students free lunches and not others so that there is a basis for comparison

Spurious Results?

- ❑ The group of students that receive free lunches must be comparable to the group that didn't
- ❑ If the two groups are inherently different, then the observed difference between them could be a result of these inherent qualities, rather than the result of receiving free lunches
 - Different starting points?
 - Different trends?
 - How do you know if the *program* caused the change in outcome, or if instead those who participate are more driven, on a particular trajectory, receiving other services, etc.?
 - Could under or over estimate impact.



Selection Bias

- ▣ Such spurious results are driven by what is known as “selection bias.”
- ▣ If the comparison group is different from the treatment group, and these differences are correlated with outcomes, then the comparison group provides a faulty estimate of the counterfactual
- ▣ How can we ensure that our comparison group accurately estimates the counterfactual?



Randomization

- ❑ Random assignment eliminates selection bias by ensuring that treatment and control subjects are identical (given a sufficient sample size). Randomization ensures a control group that accurately estimates the counterfactual....
- ❑ If we randomly assign half of all schools to receive free lunches, and the other half as “control schools”, then *on average* student ability, teacher quality, school resources, and all other school characteristics that affect test scores will be identical across the two groups
- ❑ For observable characteristics, this can be explicitly verified using simple statistical methods.
- ❑ Random assignment ensures that the only systematic difference between treatment and control groups is exposure to the treatment, and nothing else (observable or unobservable).
- ❑ Random assignment is the only “foolproof” way to establish causal impacts of a particular treatment.

2 “Counterfeit” Counterfactuals

1. Before and after:
 - Same individual before the treatment
2. Those not enrolled
 - Those who choose not to enroll in program
 - Those who were not offered the program

□ Problem:

Cannot completely know why the treated are treated and the others not

1. Before and After Examples

- Agricultural assistance program
 - Financial assistance to purchase inputs
 - Compare rice yields before and after
 - Before is normal rainfall, but after is drought
 - Find fall in rice yield
 - Did the program fail?
 - Could not separate (identify) effect of financial assistance program from effect of rainfall
- School scholarship program on enrollment

2. Non-participants Example 1

- Job training program offered
- Compare employment & earning of those who sign up to those who did not
- Who signs up?
 - Those who are most likely to benefit, i.e. those with more ability
 - Would have higher earnings than non-participants without job training
- Poor estimate of counterfactual

2. Non-participants Example 2

- Health insurance offered
- Compare health care utilization of those who got insurance to those who did not
 - Who buys insurance: those that expect large medical expenditures
 - Who does not: those who are healthy
- With no insurance: Those that did not buy have lower medical costs than that did
- Poor estimate of counterfactual

What's wrong?

- ❑ Selection bias: People choose to participate for specific reasons
- ❑ Many times reasons are related to the outcome of interest
 - Job Training: ability and earning
 - Health Insurance: health status and medical expenditures
- ❑ Cannot separately identify impact of the program from these other factors/reasons

Program placement example

- Gov't offers family planning program to villages with high fertility
- Compare fertility in villages offered program to fertility in other villages
- Program targeted based on fertility, so
 - Treatments have high fertility
 - Counterfactuals have low fertility
- Estimated program impact confounded with targeting criteria

Other Possible Solutions...

- ❑ Random assignment is not feasible in certain circumstances. In these cases, how do we assess program impacts?
- ❑ In this workshop we will also consider
 - Quasi-experiments
 - ❑ Regression discontinuity
 - ❑ Double differences
 - Instrumental Variables
- ❑ These methods attempt to estimate a counterfactual without random assignment by exploiting variation in treatment status that is not correlated with outcomes of interest (known as *exogenous* variation)

Six Common Concerns

- Quantitative versus qualitative
 - Apples and oranges.
 - RCT's can use qualitative measurement
 - Some of the most exciting work uses RCT with qualitative measurement
- Is it fair?
 - Rare (ever?) is a case of no resource constraints
 - Often random assignment is *more* fair.
- Need to know *why* something works
 - Absolutely. Helps with external validity.
 - RCT no better or worse at this.
 - Critical to understand.
 - Need theory
 - Replication helps too.

Six Common Concerns

- Spillovers
 - Will control group individuals be treated *indirectly*?
 - If so, well designed studies measure this.
- Dynamic versus static interventions
 - RCTs can handle *dynamic* too. The *process* is the intervention.
- Take too long
 - Only true *after* the program starts.
 - If planned ahead, takes no more time

Road map: The next 5 days

- Today: The Context
 - Why do results matter?
 - Linking monitoring with evaluation
 - Importance of evidence for policy
- Today, Monday, Tuesday: The Tools
 - Identification strategies
 - Sample size and power
 - Operational issues
- Wednesday, Thursday: The Experience
 - Group work on evaluation design and presentations

THANK YOU