FIGHT POVERTY WITH EVIDENCE-BASED ACTION

Introduction to Impact Evaluation

Madeleen Husselman
Country Director
IPA Ghana
THE PROBLEM

Limited Evidence of what works best to help the poor
Limited Use of available evidence

Ineffective Programs & Policies

Wasted money, enduring poverty
THE SOLUTION

- Design & Evaluate potential solutions to poverty problems
- Mobilize & Support decisionmakers to use evidence

Better Programs & Policies

MORE EVIDENCE, LESS POVERTY
IPA’s Approach

Understand market failures

Develop innovations

Use frontier knowledge from economics, political science and psychology

Impact evaluations (Randomized control trials)

Product design tests

Operational research

Test in multiple contexts

Learn when to do what

Conferences, global and local

Workshops with policy makers and practitioners

Policy memos and practitioner briefs

Direct implementation

Hands-on technical assistance

Practitioners’ toolkits
IPA EVALUATES AND COMMUNICATES WHAT WORKS

- 200 completed projects
- 235 active projects
- 250+ leading academics
- 400+ partner organizations
IPA Ghana

- In Ghana since 2005
- 17 active projects, 50+ full time staff, 2 permanent offices
- Implementation and Evaluation Teams
- Dissemination of results to local policy makers and practitioners
- Scaling ups and Replications
IPA Ghana Partners

Ghana Education Service
The World Bank
Ministry of Education
Ministry of Food and Agriculture
Sabre Education
CARE
UNICEF
Fidelity Bank
National Board for Small Scale Industries
Sinapi Aba Trust
About this Training: Objectives

- To equip participants to understand basic concepts underlying Impact Evaluations
- To increase understanding of various Evaluation Methods
- To build capacity on the application of key Theory of Change concepts
Training Agenda

• Introduction to IPA
• What is Evaluation
• Theory of Change
• What Questions Can Be Answered Through Rigorous Impact Evaluation?
• How to Randomize
• Presentations of Group Work
• Why Choose an RCT?
What is Evaluation?

Madeleen Husselman
IPA Ghana

Introduction to Impact Evaluation
Accra, October 31, 2018
Overview

1. Why Evaluate?
2. What is Evaluation?
3. Levels of Program Evaluation
4. When to do a Randomized Evaluation
5. Impact Evaluation Methods
How can impact evaluations help?

- Surprisingly little hard evidence on what works
- Can do more with given budget with better evidence
- Instead of asking “do development programs work?” we should be asking:
  - Which work best, why and when?
  - How can we scale up what works?

Think R&D, rather than ‘M&E’
Objectives of evaluation

- Accountability
  - Did we do what we said we were going to do?
  - Did we have a positive impact on people’s lives?

- Lesson learning
  - Particular programs do or do not work
  - What is the most effective route to achieve a certain outcome?
  - Similarities in strategies that are successful, for example, in changing behavior, even across fields? Underlying principles?

- Reduced poverty through more effective programs

Different types of evaluation contribute to these different objectives of evaluation
What makes a good evaluation?

- Ask the right questions
- Answers those questions in unbiased and definitive way
- Shares answers regardless of the results

To do that you need a model: a **logical framework/theory of change**
- Who is the target?
- What are their needs?
- What is the program seeking to change?
- What is the precise program or part of program being evaluated?
Monitoring and Evaluation

Evaluation

Program Evaluation

Impact Evaluation

Monitoring
What kind of evaluation should be done?

Think about who and what the evaluation is for...

- Academics
- Donors
  - Their Constituents
- Politicians / policymakers
- Technocrats
- Implementers
- Proponents, Skeptics
- Beneficiaries
Programs and their evaluations: Where do we start?

**Intervention**
- Start with a problem
- Verify that the problem actually exists
- Generate a theory of why the problem exists
- Design the program
- Think about whether the solution is cost effective

**Program Evaluation**
- Start with a question
- Verify the question hasn’t been answered
- State a hypothesis
- Design the evaluation
- Determine whether the value of the answer is worth the cost of the evaluation
## Components of Program Evaluation

<table>
<thead>
<tr>
<th>Needs Assessment</th>
<th>What is the problem?</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is the problem?</td>
<td>How, in theory, does the program fix the problem?</td>
</tr>
<tr>
<td>How, in theory, does the program fix the problem?</td>
<td>Does the program work as planned?</td>
</tr>
<tr>
<td>Does the program work as planned?</td>
<td>Were its goals achieved? The magnitude?</td>
</tr>
<tr>
<td>Were its goals achieved? The magnitude?</td>
<td>Given magnitude and cost, how does it compare to alternatives?</td>
</tr>
</tbody>
</table>

- Needs Assessment
- Program Theory Assessment
- Process Evaluation
- Impact Evaluation
- Cost Effectiveness
Levels of Program Evaluation

- Needs Assessment
- Program Theory Assessment
- Process evaluation
- Impact evaluation
- Cost-benefit / Cost-effectiveness analysis
Levels of Program Evaluation

Needs Assessment

Program Theory Assessment

Process evaluation

Impact evaluation

Cost-benefit / Cost-effectiveness analysis
Needs Assessment

What are the problems, and what are the options for addressing them?

- Target population and funding opportunities
- Need program will fill
- Articulation of program benefits
- Alternatives
- What you can change

Low learning levels for children in the Busia District in Kenya
Low Income Households
Inability to purchase books
Levels of Program Evaluation

- Needs Assessment
- Program Theory Assessment
- Process evaluation
- Impact evaluation
- Cost-benefit / Cost-effectiveness analysis
How will the program address the needs put forth in your needs assessment?

- What are the prerequisites to meet the needs?
- How and why are those requirements currently lacking or failing?
- How does the program intend to target or circumvent shortcomings?
- What services will be offered?
Levels of Program Evaluation

- Needs Assessment
- Program Theory Assessment
- Process evaluation
- Impact evaluation
- Cost-benefit / Cost-effectiveness analysis
Process Evaluation

Is your program working as planned, as described in your program theory assessment?

- Are basic tasks being completed?
- Are the services being delivered?
- Is the intervention reaching the target population?
- Is the intervention being completed well or efficiently and to the beneficiaries’ satisfaction?
Levels of Program Evaluation

- Needs Assessment
- Program Theory Assessment
- Process evaluation
- Impact evaluation
- Cost-benefit / Cost-effectiveness analysis
Impact: What is it?

How much better off students are because of the textbook distribution program than they would have been without the program.
How to measure impact?

*Impact* is defined as a comparison between:

1. The outcome some time after the program has been introduced
2. The outcome at that same point in time had the program not been introduced (the "counterfactual")
Counterfactual

- The *counterfactual* represents the state of the world that program participants would have experienced in the absence of the program (i.e. had they not participated in the program)

- **Problem**: Counterfactual cannot be observed

- **Solution**: We need to “mimic” or construct the counterfactual
Impact: What is it?

A. Positive
B. Negative
C. No impact
D. Don’t Know
Impact: What is it?

- Primary Outcome
- Time
- Intervention
- Counterfactual
- Impact
Impact: What is it?

- Primary Outcome
- Time
- Intervention
- Counterfactual
- Impact
Constructing the Counterfactual

- Counterfactual is often constructed by selecting a group not affected by the program

- Randomized:
  - Use random assignment of the program to create a control group which mimics the counterfactual.

- Non-randomized:
  - Argue that a certain excluded group mimics the counterfactual.
Non-random Treatment and Comparison Groups
What are some factors related to geographic location that are also related to your intervention?

- Proximity to schools and/or government/public offices and facilities
- Clusters based on community or religion
- Income or socio-economic status
Non-random Treatment and Comparison Groups
Types of Impact Evaluation Methods

1. Non- or Quasi-Experimental Methods
   a. Pre-Post
   b. Simple Difference
   c. Differences-in-Differences
   d. Multivariate Regression
   e. Statistical Matching
   f. Interrupted Time Series
   g. Instrumental Variables
   h. Regression Discontinuity
2. Randomized Evaluations
   Also known as:
   – Randomized Experiments
   – Random Assignment Studies
   – Randomized Field Trials
   – Social Experiments
   – Randomized Controlled Trials (RCTs)
   – Randomized Controlled Experiments
Randomly sample from area of interest
Randomly sample from area of interest

Randomly assign to treatment and control

Randomly sample from both treatment and control
Levels of Program Evaluation

- Needs Assessment
- Program Theory Assessment
- Process evaluation
- Impact evaluation

Cost-benefit / Cost-effectiveness analysis
Cost effectiveness analysis

How much does it cost to achieve specific outcomes?

- Keep a child in class for 1 extra year
- Keep a child alive for 1 extra year
- Increase yields by 20%
- Reduce CO2 emissions by 10%

This allows you to compare programs and decide which provides best value for money
Impact Evaluation Methods

Non-experimental and experimental
Impact Evaluation Methods

1. Non- or Quasi-Experimental Methods

- Pre-Post
- Simple Difference
- Differences-in-Differences
- Multivariate Regression
- Statistical Matching
- Instrumental Variables
- Regression Discontinuity
Impact Evaluation Methods

2. Randomized Experiments

Also known as:

- Random Assignment Studies
- Randomized Field Trials
- Social Experiments
- Randomized Controlled Trials (RCTs)
- Randomized Controlled Experiments
Sometimes, all we have are non-experimental methods

One end of the spectrum...
NON-EXPERIMENTAL METHODS
FOUR EXAMPLES

1. Pre-post
2. Simple difference
3. Difference-in-difference
4. Regression analysis
J-PAL conducts a test at the end of Pratham’s Balsakhi Program

Balsakhi students score an average of 51%

What can we conclude?
METHOD 1
PRE VS. POST

Look at average change in test scores over the school year for the Balsakhi children
METHOD 1
PRE VS. POST

Average test scores of Balsakhi students

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Start of program</td>
<td>24.8</td>
</tr>
<tr>
<td>End of program</td>
<td>51.22</td>
</tr>
</tbody>
</table>

Average post-test score for children with a Balsakhi: 51.22
Average pretest score for children with a Balsakhi: 24.80
Difference: 26.42
METHOD 1
PRE VS. POST - DISCUSSION

What represents the counterfactual?

How do we estimate impact?

What are the limitations of this method?
**METHOD 1**
**PRE VS. POST - DISCUSSION**

**What represents the counterfactual?**
Balsakhi students themselves—before participating in the Balsakhi program

**How do we estimate impact?**
Test scores for Balsakhi students after the program
— test scores for Balsakhi students before the program

**What are the limitations of this method?**
Assumes that the Balsakhi program was the only factor influencing any changes in test scores over time
Does not take changes in test scores over time that would anyways have occurred in the absence of the program
Divide the population into two groups:

One group not enrolled in Balsakhi program (Control)

One group enrolled in Balsakhi program (Treatment)

Compare test score of these two groups at the end of the program.
Under what conditions can the difference of -5.05 be interpreted as the impact of the Balsakhi program?

| Average score for children with a balsakhi | 51.22 |
| Average score for children without a balsakhi | 56.27 |
| Difference | -5.05 |
METHOD 2
SIMPLE DIFFERENCE - DISCUSSION

What represents the counterfactual?

How do we estimate impact?

What are the limitations of this method?
What represents the counterfactual?
Students who did not receive the Balsakhi program

How do we estimate impact?
Test scores for Balsakhi students after the program
- test scores for non-Balsakhi students after the program

What are the limitations of this method?
Assumes that Balsakhi students and non-Balsakhi students have the same characteristics; i.e., non-Balsakhi students would have had the same outcomes as those in the Balsakhi program if they had participated in the program.

There is usually something unique about the group that received the program which is WHY they received the program and others didn’t
Divide the population into two groups
- One group enrolled in Balsakhi program (Beneficiaries)
- One group not enrolled in Balsakhi program (Non-beneficiaries)

Compare the change in test scores between Treatment and Control
- i.e., difference in differences in test scores

Same thing: compare difference in test scores at post-test with difference in test scores at pretest
METHOD 3
DIFFERENCE-IN-DIFFERENCE

Average test scores

<table>
<thead>
<tr>
<th></th>
<th>Start of program</th>
<th>End of program</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enrolled in Balsakhi program</td>
<td>24.8 36.67</td>
<td>51.22 56.27</td>
</tr>
<tr>
<td>Not enrolled in Balsakhi program</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
METHOD 3
DIFFERENCE-IN-DIFFERENCE

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Post-test</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average score for children with a balsakhi</td>
<td>24.80</td>
<td>51.22</td>
<td>26.42</td>
</tr>
<tr>
<td>Average score for children without a Balsakhi</td>
<td>36.67</td>
<td>56.27</td>
<td>19.60</td>
</tr>
</tbody>
</table>
METHOD 3
DIFFERENCE-IN-DIFFERENCE

Test scores

2002 2003 Year

Non-Balsakhi students
Balsakhi students

26.42 19.60
**METHOD 3**  
**DIFFERENCE-IN-DIFFERENCE**

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Post-test</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average score for children with a Balsakhi</td>
<td>24.80</td>
<td>51.22</td>
<td>26.42</td>
</tr>
<tr>
<td>Average score for children without a Balsakhi</td>
<td>36.67</td>
<td>56.27</td>
<td>19.60</td>
</tr>
<tr>
<td>Difference</td>
<td></td>
<td></td>
<td>6.82</td>
</tr>
</tbody>
</table>

**QUESTION**: Under what conditions can 6.82 be interpreted as the impact of the balsakhi program?
METHOD 3
DIFFERENCE-IN-DIFFERENCE

What represents the counterfactual?

How do we estimate impact?

What are the limitations of this method?
METHOD 3
DIFFERENCE-IN-DIFFERENCE

What represents the counterfactual?
The non-Balsakhi students (controlling for baseline differences)

How do we estimate impact?
The change in test scores over time of Balsakhi students — changes over time of non-Balsakhi students

What are the limitations of this method?
Assumes that if the program didn’t exist, the two groups would have had identical trajectories over this period.
METHOD 4
REGRESSION ANALYSIS
METHOD 4
REGRESSION ANALYSIS

Divide the population into two groups:
One group enrolled in Balsakhi program
One group not enrolled in Balsakhi program

Compare test score of these two groups at the start and at the end of the program.

Control for additional variables like gender, class-size, income:

Post-test = $\beta_0 + \beta_1 Pre-test + \beta_2 Gender + \beta_3 Class-size + \beta_4 Balsakhi + e$
**METHOD 4**

**REGRESSION ANALYSIS**

* Significant at 5% level

<table>
<thead>
<tr>
<th>Method</th>
<th>Impact Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Pre-post</td>
<td>26.42*</td>
</tr>
<tr>
<td>(2) Simple Difference</td>
<td>-5.05*</td>
</tr>
<tr>
<td>(3) Difference-in-Difference</td>
<td>6.82*</td>
</tr>
<tr>
<td>(4) Regression with controls</td>
<td>1.92</td>
</tr>
</tbody>
</table>
Which of these methods do you think is closest to the truth?

A  Pre-post

B  Simple difference

C  Difference-in-difference

D  Regression

E  Don’t know

<table>
<thead>
<tr>
<th>Method</th>
<th>Impact Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Pre-post</td>
<td>26.42*</td>
</tr>
<tr>
<td>(2) Simple Difference</td>
<td>-5.05*</td>
</tr>
<tr>
<td>(3) Difference-in-Difference</td>
<td>6.82*</td>
</tr>
<tr>
<td>(4) Regression</td>
<td>1.92</td>
</tr>
</tbody>
</table>

*: Statistically significant at the 5% level
A randomized experiment

The other end of the spectrum...
The Basics

Take a sample of program applicants and randomly assign them to either:

**Treatment group** – is offered treatment

**Control group** – does not receive treatment (during the evaluation period)
Individuals, schools, villages, or districts are randomly selected to receive the treatment and those not selected serve as a comparison.

Two groups continue to be identical, except for treatment. Later, compare outcomes (health, test scores, etc.) between the two groups. Any differences between the groups can be attributed to the program.
Random Sampling and Random Assignment

Randomly sample from area of interest

Randomly assign to treatment and control

Randomly sample from both treatment and control
Key Advantage

Because members of the groups (treatment and control) do not differ systematically at the outset of the experiment, any difference that subsequently arises between them can be attributed to the program rather than to other factors.
RANDOMIZED EVALUATION
THE BASIC SETUP

Total Population → Target Population → Not in evaluation → Evaluation Sample → Random Assignment → Treatment Group

Control Group
# Impact of Balsakhi - Summary

<table>
<thead>
<tr>
<th>Method</th>
<th>Impact Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Pre-post</td>
<td>26.42*</td>
</tr>
<tr>
<td>(2) Simple Difference</td>
<td>-5.05*</td>
</tr>
<tr>
<td>(3) Difference-in-Difference</td>
<td>6.82*</td>
</tr>
<tr>
<td>(4) Regression</td>
<td>1.92</td>
</tr>
<tr>
<td>(5) Randomized Experiment</td>
<td>5.87*</td>
</tr>
</tbody>
</table>

*: Statistically significant at the 5% level
What if we can’t randomize?

There are more sophisticated, non-experimental methods to estimate program impacts:

• Statistical matching
• Regression discontinuity design (RDD)
• Instrumental variables
• Interrupted time series

**Common thread:** all try to mimic the counterfactual to estimate impact.

**Problem:** assumptions are not testable
## In Summary

<table>
<thead>
<tr>
<th>METHOD</th>
<th>COMPARISON GROUP</th>
<th>WORKS IF…</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Post</td>
<td>Program participants before program</td>
<td>The program was the only factor influencing any changes in the measured outcome over time</td>
</tr>
<tr>
<td>Simple Difference</td>
<td>Individuals who did not participate (data collected after program)</td>
<td>Non-participants are identical to participants except for program participation, and were equally likely to enter program before it started.</td>
</tr>
<tr>
<td>Differences in Differences</td>
<td>Same as above, plus: data collected before and after</td>
<td>If the program didn’t exist, the two groups would have had identical trajectories over this period.</td>
</tr>
<tr>
<td>Multivariate Regression</td>
<td>Same as above plus: Also have additional “explanatory” variables</td>
<td>Omitted (because not measured or not observed) variables do not bias the results because they are either: uncorrelated with the outcome, or do not differ between participants and non-participants</td>
</tr>
<tr>
<td>Propensity Score Matching</td>
<td>Non-participants who have mix of characteristics which predict that they would be as likely to participate as participants</td>
<td>Same as above</td>
</tr>
<tr>
<td>Randomized Evaluation</td>
<td>Participants randomly assigned to control group</td>
<td>Randomization “works” – the two groups are statistically identical on observed and unobserved characteristics</td>
</tr>
</tbody>
</table>
WHY RANDOMIZE?

CONCLUSION

- There are many ways to estimate a program’s impact
- Different methods can generate different estimates
- Each evaluation method has specific assumptions and limitations

If applicable, randomized experiments, when properly designed and conducted, provide the most credible method to estimate the impact of a program
More Evidence, Less Poverty

Innovations for Poverty Action
www.poverty-action.org

THANK YOU!