Measuring Impact
Outcomes & Indicators

Marc Shotland
Course Overview

1. What is evaluation?
2. Measuring impacts (outcomes, indicators)
3. Why randomize?
4. How to randomize?
5. Sampling and sample size
6. Threats and Analysis
7. RCT: Start to Finish
8. Cost Effectiveness Analysis and Scaling Up
Lecture Overview

• Theory of Change: What Do You Want to Measure?

• Theory of Measurement: What Makes a Measure Good?

• Practice of Measurement: Measuring the Unmeasurable

• Collecting Data
WHAT DO YOU WANT TO MEASURE?
Theory of change

Poor learning levels in primary school

Top-down monitoring programme

Officials receive tools and information

Intensity and frequency of monitoring increases

Teacher performance improves

Learning outcomes improve

Needs assessment

Intervention

Assumptions

Output indicators

Intermediary outcomes

Primary outcome
Theory of change: indicators

Top-down monitoring programme

- Officials receive tools & information
  - Self-reported receipt and usage rates
- Intensity and frequency of monitoring increases
  - No. of visits to schools, allocation of time & budget
- Teacher performance improves
  - Attendance, lesson plans, frequency & quality of evaluations
- Learning outcomes improve
  - Student attendance, test scores
Theory of change

Poor learning levels in primary school

Community-based monitoring and accountability programme

Parents attend meetings

Parents hold schools accountable

Teacher performance improves

Learning outcomes improve

Needs assessment

Intervention

Assumptions

Output indicators

Intermediary outcomes

Primary outcome
Theory of change: indicators

Community-based monitoring and accountability programme

Parents attend meetings

Parents hold schools accountable

Teacher performance improves

Learning outcomes improve

Indicators

- No. of meetings, Attendance
- Participation of parents, Interaction between parents and teachers
- Attendance, lesson plans, frequency & quality of evaluations
- Student attendance, test scores
Why do you need data? Follow Theory of Change

• Our theory and hypothesis helps us define the set of outcomes
• Need to find indicators that map the outcomes well
  – Characteristics: Who are the people the program works with, and what is their environment? (Sub-groups, covariates, predictors of compliance)
  – Channels: How does the program work, or fail to work?
  – Outcomes: What is the purpose of the program? Did it achieve that purpose?
WHAT MAKES A MEASURE GOOD

Theory of Measurement
The Main Challenge in Measurement:
Getting Accuracy and Precision
More accurate → More precise
Terms “Biased” and “Unbiased” Used to Describe Accuracy

More *accurate*

“Biased”
On average, we get the wrong answer

“Unbiased”
On average, we get the right answer
Terms “Noisy” and “Precise” Used to Describe Precision

More accurate

“Noisy”
Random error makes answer imprecise

“Precise”
Little random error gives precise answer

More precise
A Noisy and a Precise Measure Can Both Be Biased

- More accurate

**“Noisy”**
Random error makes answer imprecise

**“Precise”**
Little random error gives precise answer
Choices in Real Measurement Often Harder

More accurate

More precise

Noisy but unbiased

Precise but biased
Is this introducing noise or bias?

A. Noise
B. Bias

A surveyor doesn’t follow the exact phrasing of the question:

Surveyor- “you feel unsecure in this neighborhood, right?”

Respondent- “well, sometimes yes and sometimes, well, no…”

Surveyor- “so, is more of a yes, right?”

Respondent- “mmm…well”

Surveyor - “ok, thanks. Next question.”

code 1=yes
Threats to Accuracy & Precision

• In theory:
  – How well does the indicator map to the outcome? (e.g. IQ test → intelligence)

• In practice: Are you getting unbiased and precise answers?
  – Noise: in surveyors (training/quality); in responses (e.g. “How much did you spend on broccoli yesterday?” as a measure of annual broccoli spending; in data entry. Sources: forgetting, fatigue, survey length
  – Social desirability bias (response bias)
  – Recall bias (e.g. may remember upsetting events better), bias in estimation
  – Question format
    • Dimension of scale
    • Framing effect / Neutrality
    • Anchoring bias
Which is worse?

A. Bias (Low Accuracy)
B. Noise (Low Precision)
C. Equally bad
D. Depends
E. Don’t know/can’t say
A biased measure will bias our impact estimates

A. True
B. False
C. Depends
D. Don’t know
Bias is uncorrelated with treatment

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Endline</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment</strong></td>
<td><img src="image1" alt="Baseline Distribution" /></td>
<td><img src="image2" alt="Endline Distribution" /></td>
<td><img src="image3" alt="Difference Distribution" /></td>
</tr>
<tr>
<td></td>
<td><img src="image4" alt="Baseline Estimates" /></td>
<td><img src="image5" alt="Endline Estimates" /></td>
<td><img src="image6" alt="Difference Estimates" /></td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td><img src="image7" alt="Baseline Distribution" /></td>
<td><img src="image8" alt="Endline Distribution" /></td>
<td><img src="image9" alt="Difference Distribution" /></td>
</tr>
<tr>
<td></td>
<td><img src="image10" alt="Baseline Estimates" /></td>
<td><img src="image11" alt="Endline Estimates" /></td>
<td><img src="image12" alt="Difference Estimates" /></td>
</tr>
</tbody>
</table>
Bias is correlated with treatment

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Endline</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment</strong></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td></td>
<td>truth estimates</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
</tr>
</tbody>
</table>
Practice of Measurement

MEASURING THE UNMEASURABLE
The Basics

- Data that should be easy:
  - age
  - # of rooms in house
  - # in hh
What is Hard to Measure?

• (1) Things people do not know very well

• (2) Things people do not want to talk about

• (3) Abstract concepts

• (4) Things that are not (always) directly observable

• (5) Things that are best directly observed
How much beer did you consume last month?

1. 0 pints
2. 1-10 pints
3. 10-20 pints
4. 20-30 pints
5. >30 pints
1. Things people do not know very well

- **What:** Anything to estimate, particularly across time. Prone to recall error and poor estimation
  - **Examples:** distance to grocery store, profit, consumption, income, plot size

- **Strategies:**
  - Consistency checks – How much did you spend in the last week on x? How much did you spend in the last 4 weeks on x?
  - Multiple measurements of same indicator – How many minutes does it take to walk to the grocery store? How many miles away is the grocery store?
How many pints of beer did you consume yesterday

1. 0
2. 1-3
3. 4-6
4. >6

89%

11%
0%
0%
What is Hard to Measure?

(1) Things people do not know very well

(2) Things people do not want to talk about

(3) Abstract concepts

(4) Things that are not (always) directly observable

(5) Things that are best directly observed
How frequently do you yell at your spouse

A. Daily
B. Several times per week
C. Once per week
D. Once per month
E. Never
2. Things people don’t want to talk about

- **What:** Anything socially “risky” or something painful
  - **Examples:** sexual activity, alcohol and drug use, domestic violence, conduct during wartime, mental health

- **Strategies:**
  - Don’t start with the hard stuff!
  - Consider asking the question in third person
  - Always ensure the comfort and privacy of respondent
How frequently does your spouse yell at you?

A. Daily
B. Several times per week
C. Once per week
D. Once per month
E. Never

Bar chart showing:
- Daily: 47%
- Several times per week: 11%
- Once per week: 16%
- Once per month: 21%
- Never: 5%
What is Hard to Measure?

(1) Things people do not know very well

(2) Things people do not want to talk about

(3) Abstract concepts

(4) Things that are not (always) directly observable

(5) Things that are best directly observed
“I feel more empowered now than last year”

A. Strongly disagree
B. Disagree
C. Neither agree nor disagree
D. Agree
E. Strongly agree
3. Abstract concepts

- **What:** Empowerment, bargaining power, social cohesion, risk aversion…

- **Strategies:** Three steps:
  - Define what you mean
  - Choose the outcome
  - Design a good question

- **Often choice between a self-reported measure and a behavioral measure – both can add value!**
“I am involved in the decision to send my child to private vs. public school”

A. Strongly disagree
B. Disagree
C. Neither agree nor disagree
D. Agree
E. Strongly agree
How "socially connected" do you feel to the other people in this room?

- You: 30%
- Everyone else in this room: 35%

Comparison:
- A: 0%
- B: 25%
- C: 25%
- D: 10%
- E: 0%
How likely are you to walk in a neighborhood you don’t know after dark?

A. Very unlikely
B. Unlikely
C. Likely
D. Very likely
What is Hard to Measure?

(1) Things people do not know very well

(2) Things people do not want to talk about

(3) Abstract concepts

(4) Things that are not (always) directly observable

(5) Things that are best directly observed
What proportion of Black Africans are denied jobs due to racial discrimination?

A. 0%
B. 1-20%
C. 21-40%
D. 41-60%
E. >60%
4. Things that aren’t directly observable

❖ **What:** corruption, fraud, discrimination...

❖ **Strategies:**

  ▪ Audit studies, e.g. Rajasthan police experiment to register cases, Delhi doctors, “mystery shoppers”
  ▪ Multiple sources of data, e.g. inputs of funds vs. outputs received by recipients
  ▪ Don’t worry – there have already been lots of clever people before you – so do literature reviews!
What is Hard to Measure?

(1) Things people do not know very well

(2) Things people do not want to talk about

(3) Abstract concepts

(4) Things that are not (always) directly observable

(5) Things that are best directly observed
5. Things that are Best Directly Observed

- **What:** Behavioral preferences, anything that is more believable when done than said

- **Strategies:**
  - Develop detailed protocols to ensure behavioral measures are collected under the same circumstances for all individuals
The Problem

• With the following question...
Outcome: Gender Bias
Question: How effective are women leaders? (ineffective, somewhat effective, effective, very...)

A. Accuracy
B. Precision
C. Both
D. Neither
Perceptions and Attitudes

• Ask directly
  – “How effective is your leader?” (ineffective, somewhat effective, effective, very…)

• Indirect approaches often have better accuracy
  – Listen to a vignette (Male v. Female)
  – Revealed preference – voting behavior
  – Implicit association tests
Implicit Association Test

• For important outcomes that are subject to social desirability bias, it may be worth trying to measure indirectly. Implicit association is one method.

• People simplify the world for efficiency
  – Use rules of thumb to draw connections
  – May not even be aware themselves
Implicit Association Test: Match on Left or Right?
Implicit Association Test: Match on Left or Right?

Congress
Implicit Association Test: Match on Left or Right?
Implicit Association Test: Match on Left or Right?

Congress
Implicit Association Test

• People simplify the world for efficiency
  – Use thumb rules to draw connections
  – May not even be aware themselves

• Actually based on response time, not accuracy
  – Are respondents faster to select “Congress” when associated with a man than a woman?
Asking Sensitive Questions: List Randomization

- Some questions are sensitive and people are hesitant to answer truthfully.

- List randomization is a way to get the answer on average without knowing confidential information on any one person.
Asking Sensitive Questions: List

Randomization

- Randomly ask part of the sample the question with / without a sensitive option
- Response only a count, not the specific options

How many of these statements are true for you?
- This morning I took a shower.
- My nearest bank branch office is within walking distance.
- I have tea every day.

How many of these statements are true for you?
- This morning I took a shower.
- My nearest bank branch office is within walking distance.
- I have tea every day.
- I use my loan for non-business expenses.
Asking Sensitive Questions: List Randomization

<table>
<thead>
<tr>
<th>Average number of true statements:</th>
<th>Average number of true statements:</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>2.8</td>
</tr>
</tbody>
</table>

2.8 – 2.1 = 0.7 full TRUE difference (on average)

70% used their loan for non-business purposes.
List Randomization Shows Big Differences in Some Real Cases

Proportion Reporting Non-Enterprise Loan Expenditures

- Household Items
- Health
- Education

(With Standard Error Bars)
COLLECTING DATA
When to Collect Data

- [Baseline]: Before the intervention
- During the intervention
- Endline: After the intervention
- [Scale-up, more intervention(s)]
Methods of Data Collection: Not Just Surveys

- Surveys - household/individual
- Administrative data
- Logs/diaries
- Qualitative – e.g. focus groups, RRA
- Games and choice problems
- Observation
- Health/Education tests and measures
Common survey modules can be adapted for a particular project

- Demographics
- Economic
  - Income, consumption, expenditure, time use
  - Yields, production, etc.
- Beliefs
  - Expectations or assumptions
  - Bargaining power, patience, risk
- Anthropometric
- Cognitive, Learning
Primary Data Collection
Considerations

• Quality
  – Reliability and validity of the data: backchecks, double entry, data cleaning

• Costs / Logistics
  – Surveyor recruitment and training
  – Field work and transport, interview time
  – Electronic vs. paper
  – Data entry, reconciliation, cleaning, etc.

• Ethics

• Human subjects, data security
Things to take away:

• Theory of change guides measurement
  – Want to measure each step

• Data collection all about trade-offs
  – Quality and cost
  – Bias (accuracy) and noise (precision)

• Creative techniques can sometimes help
  – Think about what outcomes are most important
Thank you!