Measurement

Samer Kherfi
Course Overview

1. What is Evaluation?
2. Outcomes, Impact, and Indicators
3. Why Randomize
4. How to Randomize
5. Threats and Analysis
6. Sampling and Sample Size
7. Project from Start to Finish
8. Generalizability
Lecture Overview

• What to Measure
  – Case study (Theory of Change)

• How to measure it (well)
  – Validity, Reliability
  – How to measure the immeasurable
  – Sources of data
  – Data collection
  – Other considerations
Theory of Change (ToC)

- ToC: A causal description of how an intervention is expected to deliver the desired results (outcomes).
- A hypothesis about how and why intervention matters
Case Study

- Early 1990s: Constitutional amendment in India
- Established a modified system of governance where the village is the basic unit of local administration, reducing the power of the state government (decentralization)
- Required one third of village council seats and village council head positions for women
- Random selection of village councils
Women as policy makers

- Multiple social and economic implications to political reservations
- Broader question: Do policies change when there are more women in government?
- Focus: Provision of health and education
- Impact on health and education conditions
Women as policy makers: Theory of Change

Low investment in education and health at village level

- Quotas
  - More female council leaders
  - Women are more empowered
  - Different public goods
  - Different health and education outcomes?

Needs assessment
Intervention
Output indicators
Intermediary outcomes
Outcomes
Women as policy makers

Low investment in education and health at village level

Quotas

More female council leaders

Women are more empowered

Village leader has decision power

Women have different preferences

Different public goods

Different health and education outcomes?

Public goods reflect women’s preferences

Needs assessment

Intervention

Assumptions

Output indicators

Intermediary outcomes

Outcomes
Women as policy makers

Low investment in education and health at village level

**Quotas**

- More female council leaders
  - Gender of leader

- Women are empowered
  - Female attendance in council meetings; Time women speak
  - Budget allocation; Type of public goods available

- Different public goods

- Different health & education outcomes?
  - Literacy levels, BMI of children
Women as policy makers

Low investment in education and health at village level

Quotas

- More female council leaders
- Women are empowered
- Different public goods
- Different health & education outcomes?

Indicators

- Gender of leader
- Female attendance in council meetings; Time women speak
- Budget allocation; Type of public goods available
- Literacy levels, BMI of children

Results

- More female leaders
- Female attendance ↑ Active participation
- Allocations shift; More investment in roads, drinking water
- Not yet measured
Log Frame

- Logical Framework: A management tool used to facilitate the design, execution, and evaluation of an intervention. It involves identifying strategic elements (inputs, outputs, outcomes and impact) and their causal relationships, indicators, and the assumptions and risks that may influence success and failure.
### Log Frame: Encouraging youth apprenticeship

<table>
<thead>
<tr>
<th>Impact (Goal/Overall objective)</th>
<th>Objectives Hierarchy</th>
<th>Indicators</th>
<th>Sources of Verification</th>
<th>Assumptions / Threats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impact (Goal/Overall objective)</td>
<td>Better labor-market situation for the youth</td>
<td>Apprenticeship rate, employment rate and qualifications</td>
<td>Administrative data, national surveys</td>
<td>Improving matching and supply side policies lead to equilibrium shift</td>
</tr>
<tr>
<td>Outcome (Project Objective)</td>
<td>Youth manage to enter (and to stay) in apprenticeship</td>
<td>Apprenticeship entry, dropout</td>
<td>Administrative data, dedicated survey</td>
<td>Counseling is efficient</td>
</tr>
<tr>
<td>Outputs</td>
<td>Youth meet the counselors, counselors track youth progress</td>
<td>Information about meetings and contacts?</td>
<td>Administrative data, survey (?)</td>
<td>Treatment is implemented in a satisfactory way</td>
</tr>
<tr>
<td>Inputs (Activities)</td>
<td>Counselors are trained and made available</td>
<td>Hours dedicated to the program</td>
<td>Agreement</td>
<td>Government believes apprenticeship should increased</td>
</tr>
</tbody>
</table>

## Data used

<table>
<thead>
<tr>
<th>Sources of Measurement</th>
<th>Indicators</th>
</tr>
</thead>
</table>
| Survey (youth)                                              | • Employment/Apprenticeship status  
• Socio-demographics  
• Intensity of the program                                   |
| Administrative data from vocational centers and counselors  | • Registration for vocational degree  
• Meetings: numbers, frequency                                   |
| Apprenticeship contracts                                    | • Employer characteristics                                                 |
| Survey (vocational centers)                                 | • Course attendance  
• Youth experienced difficulties?                                 |
| Other (existing) surveys                                    | • Contextual variables  
• Outcomes about youth in the same area                             |
### Results

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Group</th>
<th>Control</th>
<th>Treated</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entered apprenticeship</td>
<td></td>
<td>51%</td>
<td>57%</td>
<td>0.06**</td>
</tr>
<tr>
<td>Dropout from apprenticeship</td>
<td></td>
<td>21%</td>
<td>25%</td>
<td>0.04**</td>
</tr>
</tbody>
</table>
HOW TO MEASURE IT (WELL)

• The basics
CENSUS EXAMPLE
The Basics

• Data that should be easy?
  – E.g. Age, # of rooms in house, # in hh

• What is the survey question identifying?
  – E.g. Are hh members people who are related to the household head? People who eat in the household? People who sleep in the household? Bobcats?
When the obvious is not so obvious…

• Let’s think about the people who eat from the same pot in the household where you usually stay. How many adults, adolescents, and children? Adults are age 18 and older, adolescents are ages 13 to 17, and children are ages 12 and younger.

— So in total there are how many people in the household where you usually stay? DON’T ADD TOTAL FOR RESPONDENT.
Validity, Reliability

- How to measure it (well)
The main challenge in measurement

- Accuracy

- Precision
The main challenge in measurement

- Validity

- Reliability
Validity

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

• In practice:
  – Are your survey questions unbiased?
  – Potential biases:
    • Social desirability bias
    • Demand bias (response bias)
    • Framing effect
    • Recall bias
    • Anchoring bias
Reliability

• In theory:
  – The measure is consistent, precise, but not necessarily valid

• In practice:
  – Length, fatigue
  – “How much did you spend on broccoli yesterday?” (as a measure of annual broccoli spending)
  – Ambiguous wording (definitions, relationships, recall period)
  – Answer choice (open/closed, Likert, ranked)
General noise

• Surveyor training/quality
• Data entry
• Poor translation
Which is worse?

A. Poor Validity
B. Poor reliability
C. Equally bad
D. Depends
E. Don’t know/can’t say
“Consistently Biased”

<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" /></td>
<td><img src="image2" alt="Endline" /></td>
<td><img src="image3" alt="Difference" /></td>
</tr>
<tr>
<td></td>
<td><img src="image4" alt="truth estimates" /></td>
<td><img src="image5" alt="red bars" /></td>
<td><img src="image6" alt="red bars" /></td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td><img src="image7" alt="Baseline" /></td>
<td><img src="image8" alt="Endline" /></td>
<td><img src="image9" alt="Difference" /></td>
</tr>
<tr>
<td></td>
<td><img src="image10" alt="red bars" /></td>
<td><img src="image11" alt="red bars" /></td>
<td><img src="image12" alt="red bars" /></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="image1" alt="Baseline" /></td>
<td><img src="image2" alt="Endline" /></td>
<td><img src="image3" alt="Difference" /></td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td><img src="image4" alt="Baseline" /></td>
<td><img src="image5" alt="Endline" /></td>
<td><img src="image6" alt="Difference" /></td>
</tr>
</tbody>
</table>

Bias is indicated by the difference between the truth and the estimates.
Measuring the immeasurable

• How to measure it (well)
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
Why the Hard to Measure?

- Missing key characteristics that:
  - interact with policies to change their impact
  - help us tailor policies and programs to better reach stated objectives

- If we can't measure it, we can't evaluate its importance
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 health center, 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 health center? How many kilometers away is the health center?
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 question in third person
- Always ensure comfort and privacy of respondent
3. Abstract concepts

**What:** Potentially the most challenging and interesting type of difficult-to-measure indicators

- **Examples:** empowerment, bargaining power, social cohesion, risk aversion

- **Strategies:**
  - Three key steps when measuring “abstract concepts”
    - Define what you mean by your abstract concept
    - Choose the outcome that you want to serve as the measurement of your concept
    - Design a good question to measure that outcome

- Often choice between choosing a self-reported measure and a behavioral measure – both can add value!
Choosing Fruit vs. Chocolate
Read and van Leeuwen (1998)

If you were deciding today, would you choose fruit or chocolate for next week?
Today, 74% of subjects choose fruit for next week.
Choosing & Eating Simultaneously

If you were deciding today, would you choose fruit or chocolate for today?
Today, 70% of subjects choose chocolate for today.
Time Inconsistent Preferences

100 dollars today or 110 dollars after 1 month? 100

100 dollars after one year or 110 dollars after 1 year and 1 month? 110
I, ____________________________, commit to save for ____________________________.

I have opened a SEED savings account with a

Goal Date / Goal Amount of ____________.

I will try everything in my power to accomplish my SEED Savings Goal by

saving _______________ Pesos a day / a week.

If I achieve this goal, I will be able to enjoy my savings to ____________________________

by ____________________________.

__________________________  ____________________________

Name                                      Date
Things that aren’t directly observable

**What:** You may want to measure outcomes that you can’t ask directly about or directly observe

- **Examples:** corruption, fraud, discrimination

**Strategies:**

- Sometimes you just have to be clever…
- Don’t worry – there have already been lots of clever people before you – so do literature reviews!
Perceptions and Attitudes

• “How effective is your leader?” (ineffective, somewhat effective, effective, very…)
  – Survey response: explicit distaste for female leaders (Feeling Thermometer)

• Listen to a Vignette (Male v. Female)
  – Bias large and significant for male listeners

• Revealed preference – voting behavior

• Implicit Association tests
  – Increased likelihood of associating women with leadership activities
Implicit Association Test
Things that are best directly observed

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

**Strategies:**

- Develop detailed protocols
- Ensure data collection of behavioral measures done under the same circumstances for all individuals
Sources of data
Where can we get data?

• **Administrative Data**
  – State government census data
  – School data
  – Cell phone usage

• **Other Secondary Data**
  – National surveys, geographic data…

• **Primary Data**
  – Your own survey
Primary Data Collection

• Self-reported Surveys
  – Phone, face-to-face, internet, cell phone?
• Exams, tests, etc
• Games
• Diaries
• …
Modules

- Income, consumption, expenditure
- Perceptions, expectations, aspirations
- Bargaining power
- Patience, risk
- Behavior (time use)
- Anthropometric
- Cognitive, Learning
- Yields
Why collect your own data?

- The standard RCT design is
  - Baseline
  - During the intervention
  - Endline
  - Scale-up, intervention

- Pros vs. cons of collecting your own data
  - Scale, cost
  - Focus of questions
Data Collection Considerations

- Quality Control
- Surveyor training
- Surveyor (gender) composition
- Human subjects
- Data Security
- Electronic v paper
- Costs
OTHER CONSIDERATIONS

• How to measure it (well)
Don’t forget

• Ethics
• Might affect compliance
• Respondent (and interviewer) fatigue
Thank you