Measurement

Jennifer Peck
Swarthmore College
Course Overview

1. What is Evaluation?
2. Outcomes, Impact, and Indicators
3. Why Randomize?
4. How to Randomize?
5. Project from Start to Finish
6. Generalizability
Lecture Overview

1. What to Measure
   - Theory of Change, Outcomes, Indicators
     • Women as Policymakers Case Study

2. How to Measure It (Well)
   - Indicators: validity and reliability
   - Hard-to-measure outcomes
   - Sources of Data
   - Collecting Data
Reforming School Monitoring

CASE STUDY
Theory of Change: Top-down Monitoring

1. Poor learning levels in primary school
2. Top-down monitoring programme
3. Officials receive tools and information
   - Use of tools increases monitoring
   - Increased monitoring causes increased teacher effort
4. Intensity and frequency of monitoring increases
5. Teacher performance improves
6. Learning outcomes improve

Secondary components:
- Needs assessment
- Intervention
- Assumptions
- Output indicators
- Intermediary outcomes
- Primary outcome
## Log Frame: Top-down Monitoring

<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>Learning outcomes improve</td>
<td>Student attendance,</td>
<td>Surveys,</td>
<td>Poor teacher performance is primary cause of poor learning outcomes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Test scores</td>
<td>Testing</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome (Project Objective)</th>
<th>Teacher performance improves</th>
<th>Attendance, Lesson plans, Frequency &amp; quality of evaluations</th>
<th>Survey, Admin data, spot checks, testing</th>
<th>Teacher effort is dependent on monitoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outputs</td>
<td>Intensity &amp; frequency of monitoring increases</td>
<td>No. of visits to schools, allocation of time and budget</td>
<td>Surveys, admin data, spot checks</td>
<td>Use of tools increases monitoring</td>
</tr>
<tr>
<td>Inputs (Activities)</td>
<td>Officials receive tools &amp; information</td>
<td>Self-reported receipt &amp; usage rates</td>
<td>Surveys</td>
<td>Sufficient materials, funding, manpower</td>
</tr>
</tbody>
</table>
Theory of Change: Community-based Monitoring

- Poor learning levels in primary school
- Community-based monitoring and accountability programme
  - Parents are interested in
  - Parents attend meetings
  - Parents hold schools accountable
  - Teacher performance improves
  - Learning outcomes improve

- Needs assessment
- Intervention
- Assumptions
- Output indicators
- Intermediary outcomes
- Primary outcome
# Log Frame: Community-based monitoring

<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>Learning outcomes improve</td>
<td>Student attendance, Test scores</td>
<td>Surveys, Testing</td>
<td>Poor teacher performance is primary cause of poor learning outcomes</td>
<td></td>
</tr>
<tr>
<td>Teacher performance improves</td>
<td>Attendance, Lesson plans, Frequency &amp; quality of evaluations</td>
<td>Survey, Admin data, spot checks, testing</td>
<td>Teacher effort is dependent on parent accountability</td>
<td></td>
</tr>
<tr>
<td>Parents holds teacher accountable</td>
<td>Participation of parents, Interaction between parents &amp; teachers</td>
<td>Surveys</td>
<td>Parents care about their children’s education</td>
<td></td>
</tr>
<tr>
<td>Meetings happen, Parents attend meetings</td>
<td>No. of meetings, attendance</td>
<td>Surveys, observations</td>
<td>Sufficient funding, manpower</td>
<td></td>
</tr>
</tbody>
</table>
Use in analysis

INDICATORS
CENSUS EXAMPLE
The Basics

- Lesson: don’t be complacent even about “easy” data
  - 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?
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. IQ tests ➔ intelligence)

• In practice: are you getting unbiased answers?
  – Social desirability bias
  – Framing effect
  – Recall bias
  – Anchoring bias
Reliability

• In theory:
  – The measure is consistent and precise vs. “noisy”

• In practice: many things can reduce reliability
  – 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)
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="#" alt="Baseline" /></td>
<td><img src="#" alt="Endline" /></td>
<td><img src="#" alt="Difference" /></td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td><img src="#" alt="Baseline" /></td>
<td><img src="#" alt="Endline" /></td>
<td><img src="#" alt="Difference" /></td>
</tr>
</tbody>
</table>

truth estimates
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></td>
<td><img src="image4" alt="Truth" /></td>
<td><img src="image5" alt="Estimates" /></td>
<td></td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td><img src="image6" alt="Baseline" /></td>
<td><img src="image7" alt="Endline" /></td>
<td><img src="image8" alt="Difference" /></td>
</tr>
<tr>
<td></td>
<td><img src="image9" alt="Truth" /></td>
<td><img src="image10" alt="Estimates" /></td>
<td></td>
</tr>
</tbody>
</table>
Things to Think About

- Question wording, definitions, recall period
- Answer choice
  - Open/closed, single v. multiple options, units
- Surveyor training/quality
- Data entry
- Length, fatigue
- Translation
  - Back-translate and pretest in local languages
The problem

• With the following questions…
Outcome: annual consumption
Indicator: food expenditure in last week

A. Validity
B. Reliability
C. Both
D. Neither
Outcome: annual consumption
Indicator: food expenditure in last three months

A. Validity
B. Reliability
C. Both
D. Neither
HARD-TO-MEASURE INDICATORS
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 money did you spend on coffee in the past two weeks?

- < $5
- $6-$10
- $11-15
- $16-$20
- >$21
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
- Get information indirectly, if possible
- List randomization
“I am a risk-taker.”

A. Strongly disagree
B. Disagree
C. Neither agree nor disagree
D. Agree
E. Strongly agree
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!
4. 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!
5. 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
Women’s Empowerment?

- Voters respect women’s voices (gender bias)
- Parents invest in daughters
- Women’s empowerment
- More female pradhans
- Women can be re-elected
- Career opportunities for women (in politics)
- Reservations for women
Perceptions and Attitudes

• “How effective is your leader?” (ineffective, somewhat effective, effective, very…)
• Listen to a Vignette (Male v. Female)
• Revealed preference – voting behavior
• Implicit Association Tests
  – https://implicit.harvard.edu
Implicit Association Test
Results on Women’s Empowerment

• Significant electoral gains for women in subsequent unreserved elections
• Changed perceptions of women’s ability to lead effectively
• Heightened career aspirations of adolescent girls and increased level of educational attainment
SOURCES OF DATA
Where can we get data?

• Administrative Data
• Other Secondary Data
• Primary Data
Primary Data Collection

- Surveys
- Exams, tests, etc
- Games
- Vignettes
- Direct Observation
- Diaries/Logs
Modules

- Income, consumption, expenditure
- Perceptions, expectations, aspirations
- Bargaining power
- Patience, risk
- Behavior (time use)
- Anthropometric
- Cognitive, Learning
- Yields
Considerations

DATA COLLECTION
Data Collection Considerations

- Quality Control
- Surveyor training
- Surveyor (gender) composition
- Human subjects
- Data Security
- Electronic vs paper
- Costs
When to collect Data

• Baseline
• During the intervention
• Endline
• Scale-up, intervention
QUESTIONS?