Measurement

Roland Rathelot

J-PAL
Course Overview

1. What is Evaluation?
2. Outcomes, Impact, and Indicators
3. Why Randomize and Common Critiques
4. How to Randomize
5. Sampling and Sample Size
6. Threats and Analysis
7. Project from Start to Finish
8. Cost-Effectiveness Analysis and Scaling Up
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
Case study

• Encouraging apprenticeship among the youth
## Log frame

<table>
<thead>
<tr>
<th></th>
<th>Objectives Hierarchy</th>
<th>Indicators</th>
<th>Sources of Verification</th>
<th>Assumptions / Threats</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Impact</strong></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><strong>Overall objective</strong></td>
<td><strong>Impact</strong> (Goal/Overall objective)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Outcome</strong></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><strong>Project Objective</strong></td>
<td><strong>Outcome</strong> (Project Objective)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Outputs</strong></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><strong>Inputs</strong></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>
<tr>
<td><strong>Activities</strong></td>
<td></td>
<td></td>
<td></td>
<td></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></th>
<th></th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>Treated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entered apprenticeship</td>
<td>51%</td>
<td>57%</td>
<td></td>
<td>0.06**</td>
</tr>
<tr>
<td>Dropout from apprenticeship</td>
<td>21%</td>
<td>25%</td>
<td></td>
<td>0.04**</td>
</tr>
</tbody>
</table>
HOW TO MEASURE IT (WELL)

• The basics
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
“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="truth-estimates.png" alt="Diagram" /></td>
<td><img src="endline.png" alt="Diagram" /></td>
<td><img src="difference.png" alt="Diagram" /></td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td><img src="baseline.png" alt="Diagram" /></td>
<td><img src="baseline.png" alt="Diagram" /></td>
<td><img src="baseline.png" alt="Diagram" /></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>
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!
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?
Time Inconsistent Preferences

Choosing & Eating Simultaneously

Today, 70% of subjects choose chocolate for today.
“...but you must bind me hard and fast, so that I cannot stir from the spot where you will stand me... and if I beg you to release me, you must tighten and add to my bonds.”

- The Odyssey
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
Taking a real IAT
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

- Despite advent of modern contraception, more than 50% of births are reported as unwanted

- Contraceptive methods often stocked out, rationed by wait times

- Official policy that husband consent is not required to obtain contraceptives; not in rural areas. Many husbands unaware of birth control use (21%)
The baby gave you nine months' notice. Where're the napkins, shawls and everything?

There's nothing at all, sister. In fact we've problems in feeding and dressing the other ten children at home.

Take off your shirt so that we can wrap in the baby.

Okay sister.

You men're to blame because you always discourage your wives to use any of the many family planning methods!

Never again darling - family planning makes sense!

Poverty, disease...
Free, Instant Access to Depo Provera or Jadelle

If you bring this voucher with your NRC card to Nurse Grace Daka at Chipata Clinic between xx and xx hours Monday through Saturday, we guarantee that you will receive:

- very quick, first-class personalized family planning services from Grace Daka, your own dedicated family planning nurse only for the lucky women in this study
- one implant of Jadelle or one years’ worth of Depo Provera if the nurse deems it medically appropriate
- a wait time of no more than 30 minutes; we will give you a free gift if you wait longer than 30 minutes
- a free, surprise gift for you, the woman of the house, if you are one of the first 50 women to see Grace Daka with this voucher

Name ........................................................................
NRC# .................................................................
Date ........................................................................

sponsored by

see inside for details
Free, Instant Access to Depo Provera or Jadelle

If you bring this voucher with your NHC card to Grace Clinic or Cipero Clinic between six and six hours Monday through Saturday, we guarantee that you will receive:

- A very quick, free, and custom-designed family planning service from Grace Clinic, your own dedicated family planning nurse for the lady women in your study
- One implant of Jadelle or one year’s worth of Depo Provera if the woman wants it medically appropriate
- A wait time of no more than 30 minutes; we will give you a free gift if your wait, longer than 30 minutes
- A free, surprise gift for you, the women of the house, if you come in as the first 50 women to see Grace Clinic with this voucher
Results

Women who received the voucher alone:

• 23% more likely to visit a family planning nurse
• 38% more likely to take up a concealable form of contraception
• 57% reduction in unwanted births

...than when they were given the voucher with their husband.

Only when women have greater autonomy to adopt contraception does it lead to a decrease in unwanted births.

...but is that the whole story?