Measurement

Rohit Naimpally
Research & Training Manager
J-PAL Global

Course Overview

1. What is Evaluation?
2. Outcomes, Impact, and Indicators
3. Why Randomize?
4. How to Randomize
5. Sampling and Sample Size
6. Threats and Analysis
7. Evaluation from Start to Finish
8. Evidence from Community-Driven Development, Health, and Education Programs
9. Using Evidence from Randomized Evaluations

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

A framing quote…

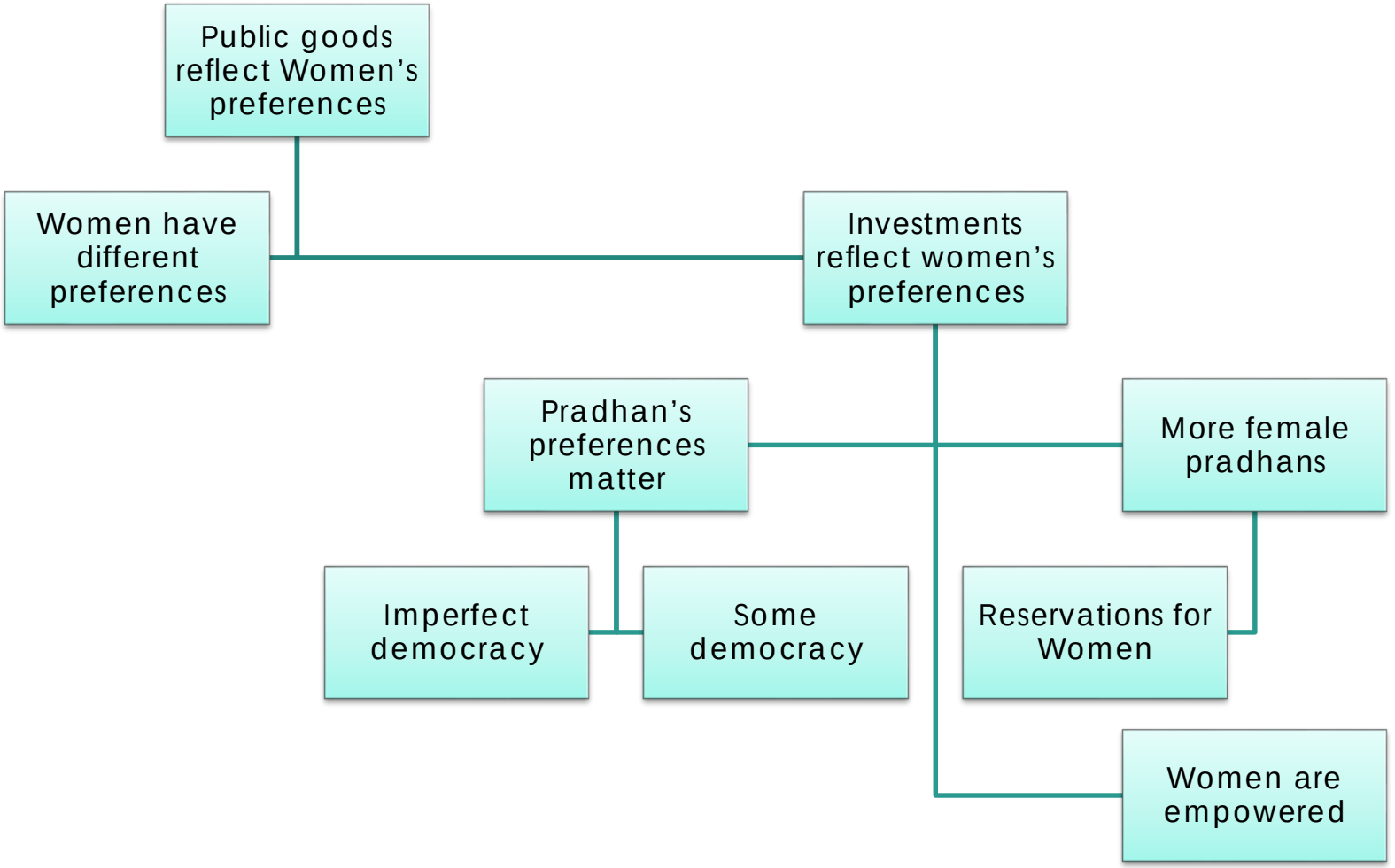
Science begins with counting. To understand a phenomenon, a scientist must first describe it; to describe it objectively, he must first measure it…measured in some reliable, reproducible way.

- Siddhartha Mukherjee, "The Emperor of All Maladies"

Women as Policymakers

CASE STUDY

Theory of Change



Log Frame

	Objectives Hierarchy	Indicators	Sources of Verification	Assumptions / Threats
Impact (Goal/ Overall objective)	Public good investment represents women's preferences			
Outcome (Project Objective)	Women voice political views			
Outputs	More female Pradhans			
Inputs (Activities)	Reservations for women			

Source: Roduner, Schlappi (2008) Logical Framework Approach and Outcome Mapping, A constructive Attempt of Synthesis,

Log Frame

	Objectives Hierarchy	Indicators	Sources of Verification	Assumptions / Threats
Impact (Goal/ Overall objective)	Public good investment represents women's preferences	Government spending	Administrative data: Budgets, Balance Sheets	Pradhan preferences matter: imperfect/some democracy
Outcome (Project Objective)	Women voice political views	Number of times a woman spoke	Transcript from village meeting	Women develop independent views
Outputs	More female Pradhans	Whether or not a Panchayat had a female Pradhan	Administrative records	The law is implemented, there is no backlash
Inputs (Activities)	Reservations for women	Law is passed	The constitution	The government realizes the need for women representation

Source: Roduner, Schlappi (2008) Logical Framework Approach and Outcome Mapping, A constructive Attempt of Synthesis,

Data used

Sources of Measurement	Indicators
Household (HH) Survey	• Declared HH preferences • HH perceptions of quality of public goods and services
Village Leader Interview	• Political experience • Investments undertaken
Village PRA	• Village infrastructure + investments • Perception of public good quality • Participation of men and women • Issues
Administrative Data	• Budgets • Balance sheets
Transcript from village meeting	• Who speaks and when (gender) • Issues raised

Results

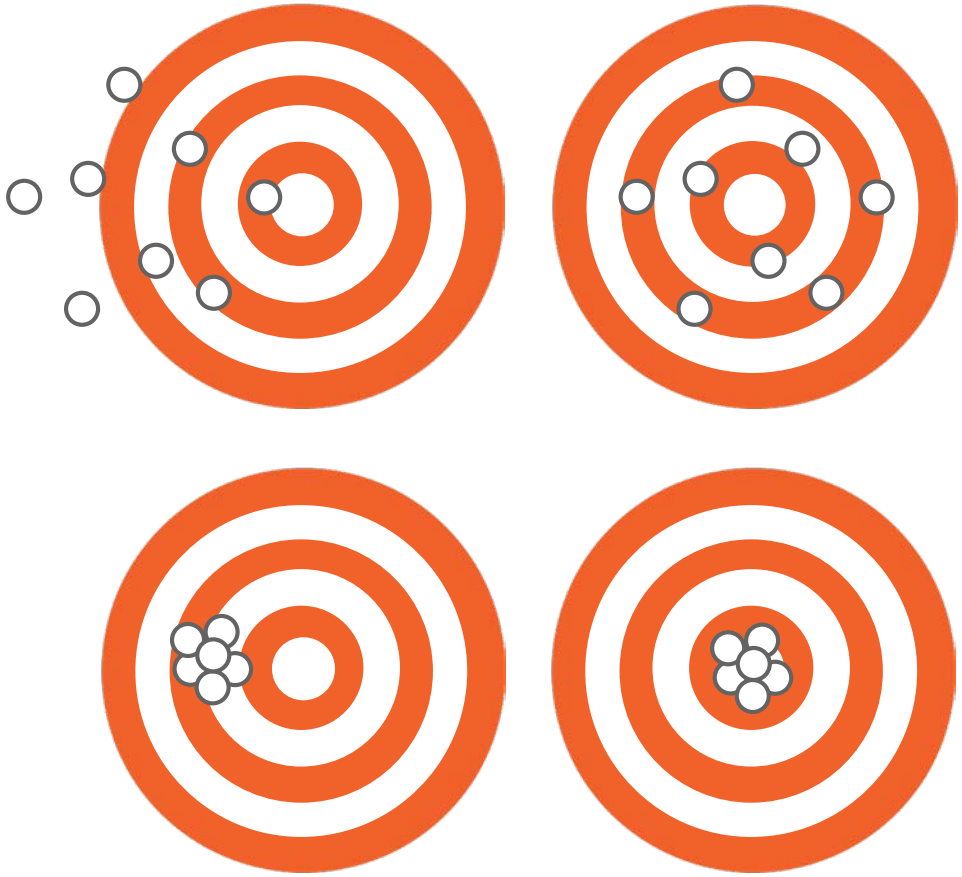
		West Bengal			Rajasthan		
		Issue			Issue		
Issue	**Investment**	**W**	**M**	**Reserved Investment**	**W**	**M**	**Reserved Investment**
Drinking Water	# facilities	**31 %**	17%	9.09*	**54%**	49%	2.62*
Road Improvement	Road Condition (0-1)	**31 %**	25%	0.18*	13%	**23%**	-0.08*
Irrigation	# facilities	4%	**20 %**	-0.38	2%	4%	-0.02
Education	Informal education center	6%	**12 %**	-0.06	5%	**13%**	-

Use in analysis

INDICATORS

The main challenge in measurement

- Accuracy
- Precision



The main challenge in measurement

- Validity

- Reliability

Validity

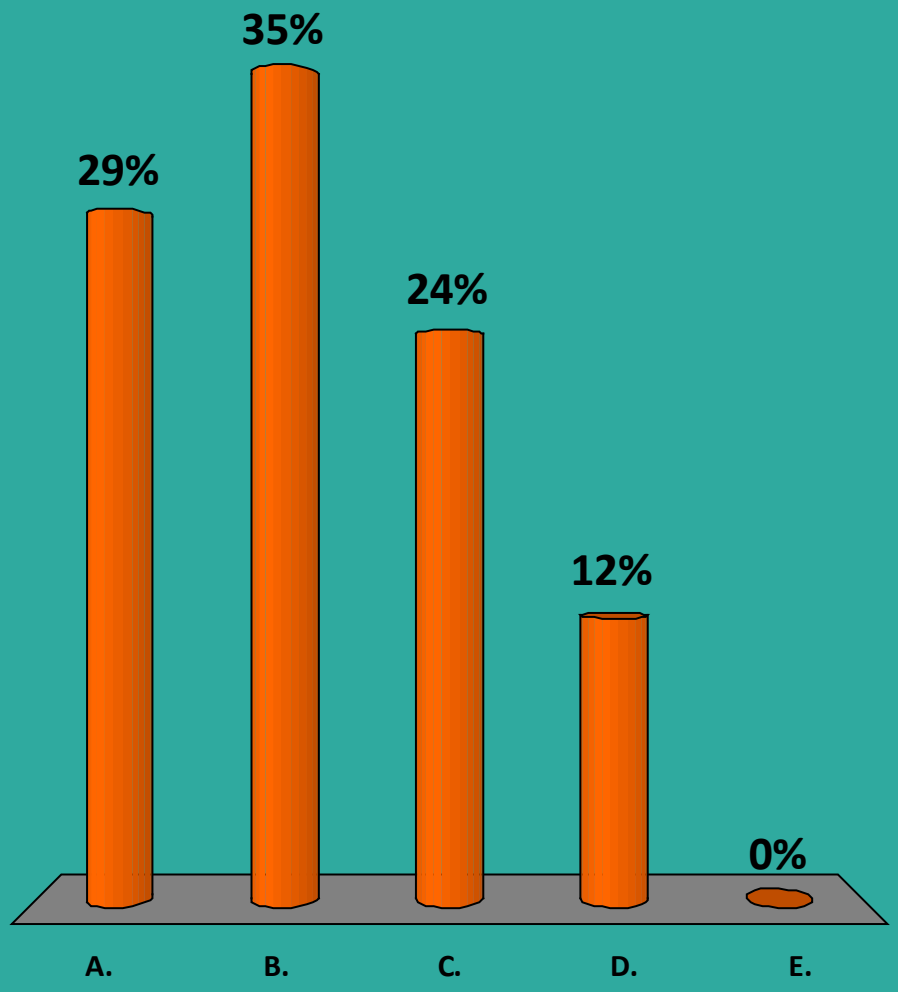
- In theory:
  - How well does the indicator map to the outcome? (e.g. IQ tests → intelligence)
- In practice: are you getting unbiased answers?
  - Social desirability bias
  - Framing effect
  - Recall bias
  - Anchoring bias

Reliability

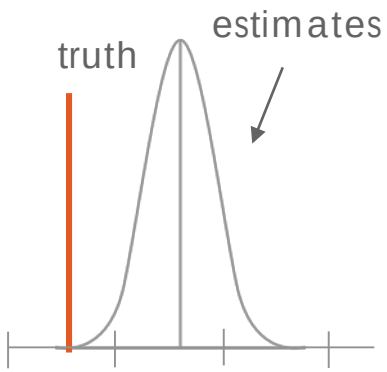
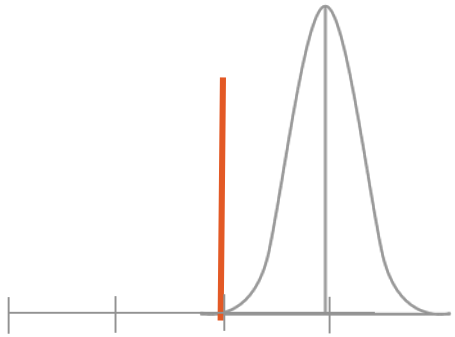
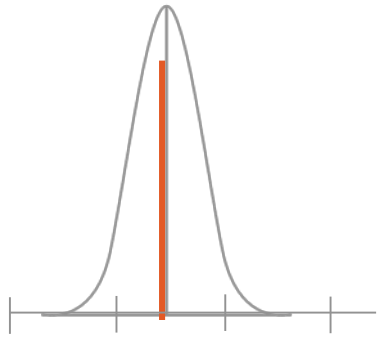
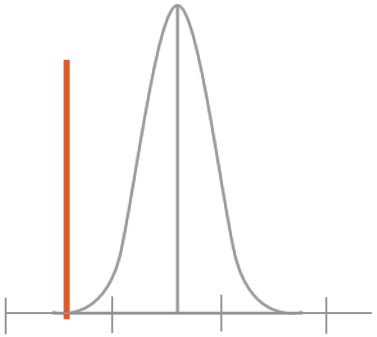
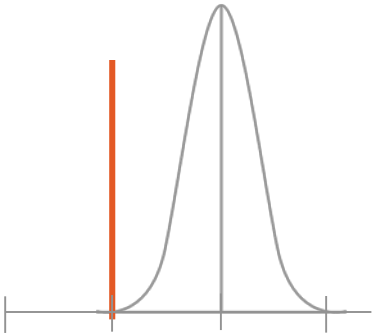
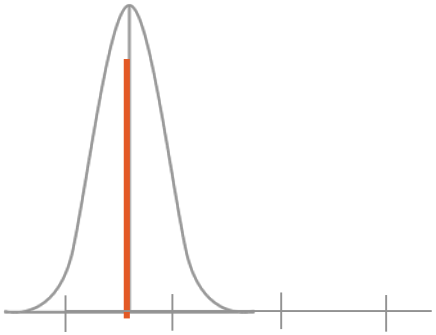
- In theory:
  - The measure is consistent and precise vs. "noisy"
- In practice: many things can reduce reliability
  - Length, fatigue
  - "How much did you spend on broccoli yesterday?" (as a measure of annual broccoli spending)
  - Ambiguous wording (definitions, relationships, recall period)
  - Answer choice (open/closed, Likert, ranked)

Which is worse?

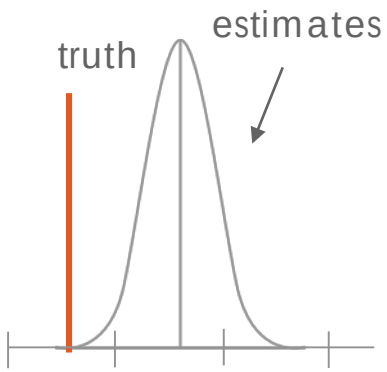
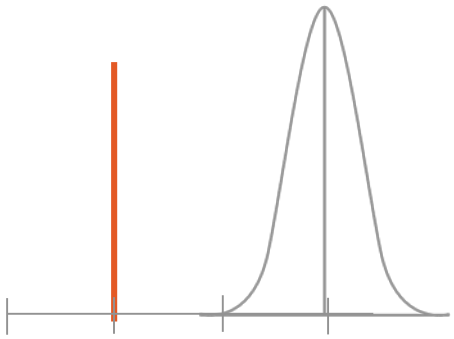
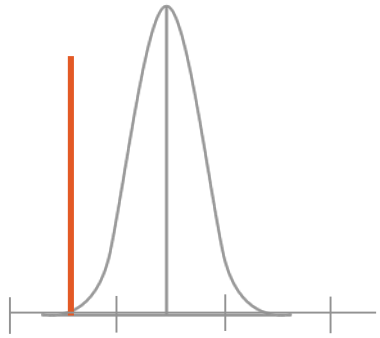
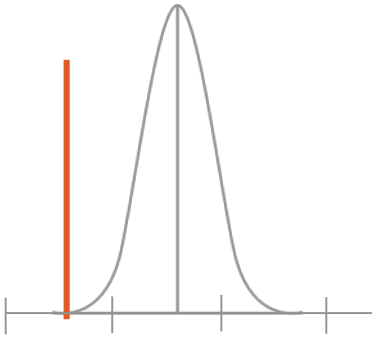
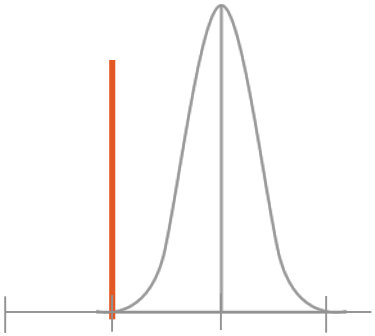
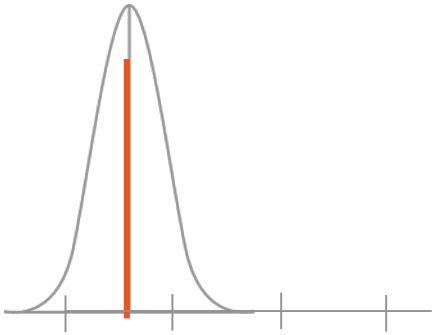
A. Poor Validity
B. Poor Reliability
C. Equally bad
D. Depends
E. Don't know/can't say



"Consistently Biased"

	Baseline	Endline	Difference
Treatment			
Control			

Bias is correlated with treatment

	Baseline	Endline	Difference
Treatment			
Control			

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…

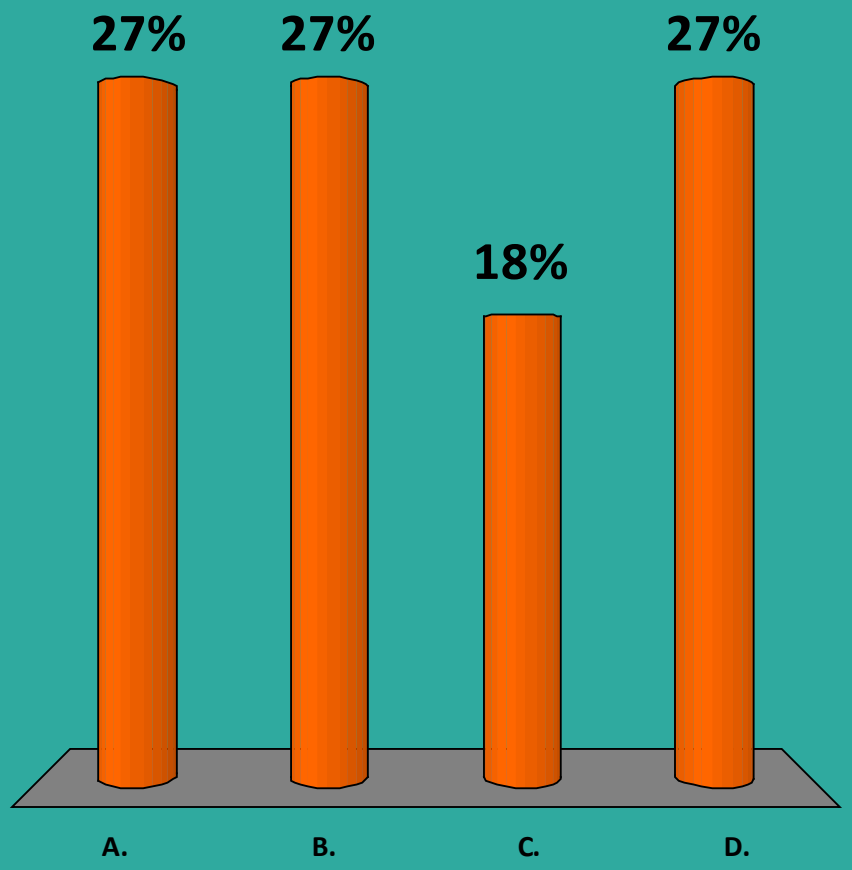
Outcome: annual consumption
Indicator: food expenditure in last week

A. Validity

B. Reliability

C. Both

D. Neither



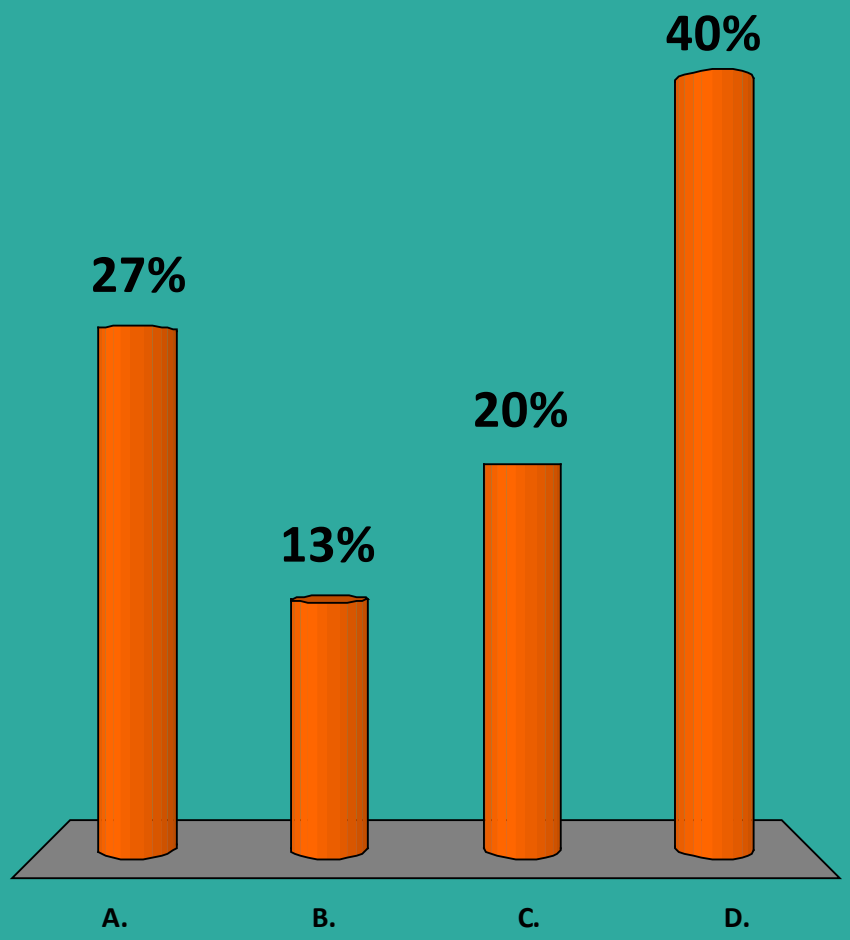
Outcome: annual consumption
Indicator: food expenditure in last three months

A. Validity

B. Reliability

C. Both

D. Neither



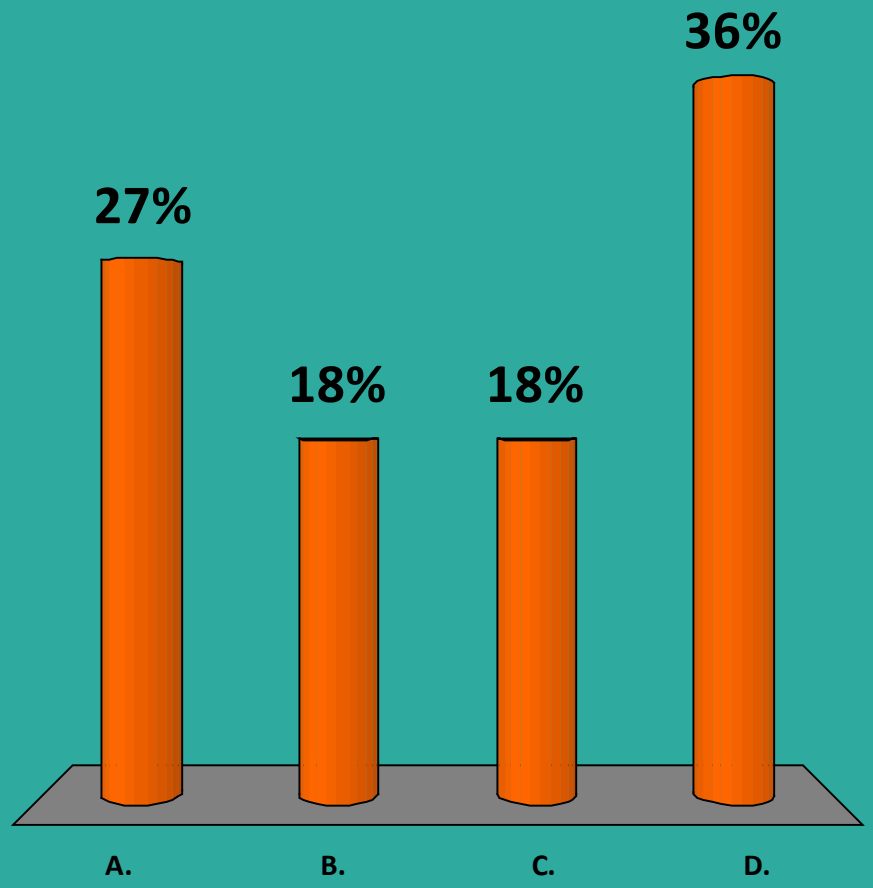
Question: have you had sex in the past week?
[if yes] Did you use protection?

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?

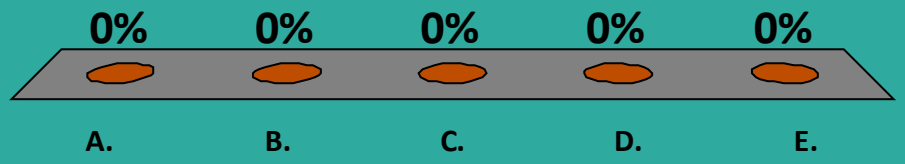
A. < $5

B. $6-$10

C. $11-15

D. $16-$20

E. >$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?

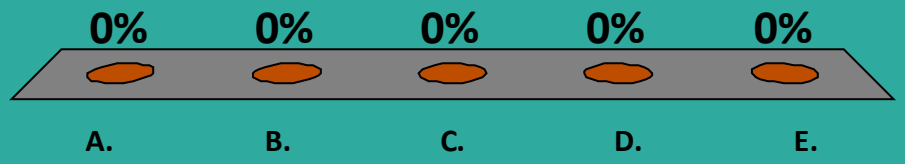
2. Things people don't want to talk about

- 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

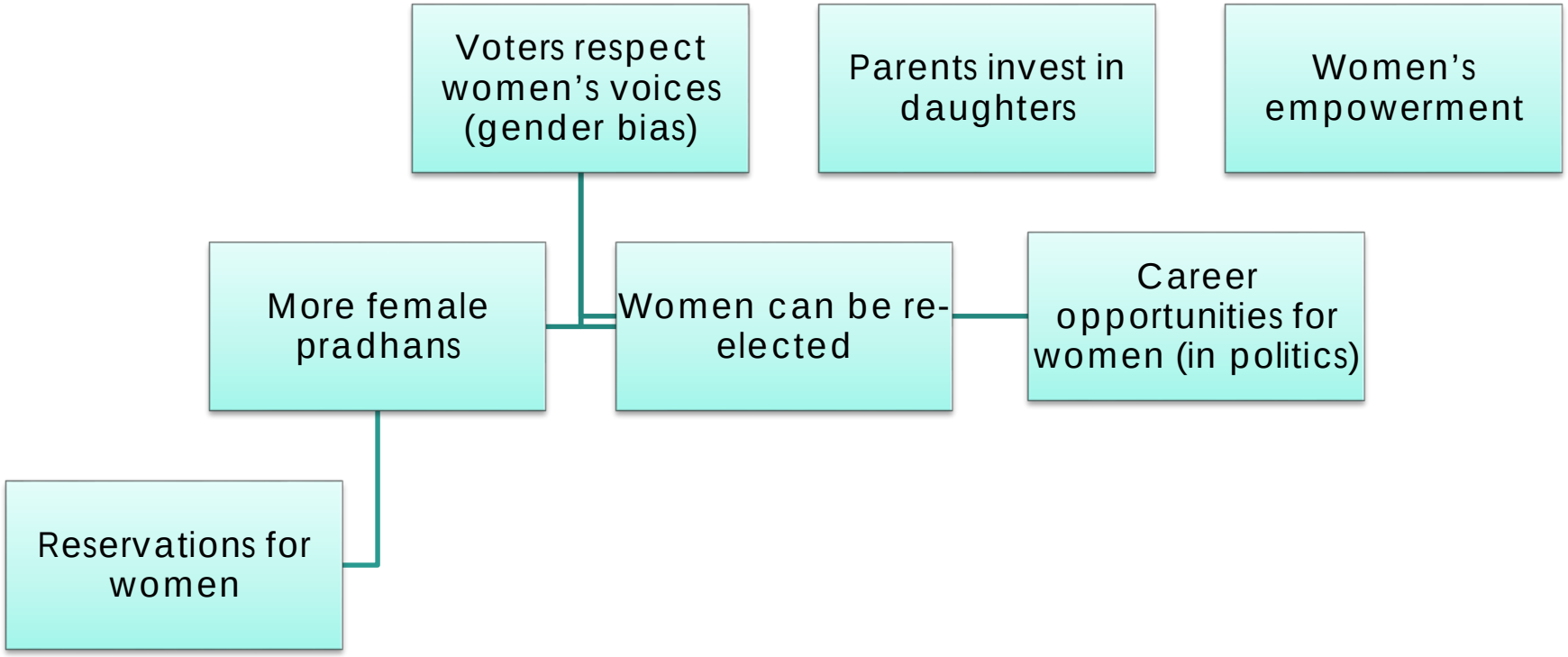
- Strategies:
  - Sometimes you just have to be clever…
  - Don't worry – there have already been lots of clever people before you – so do literature reviews!

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

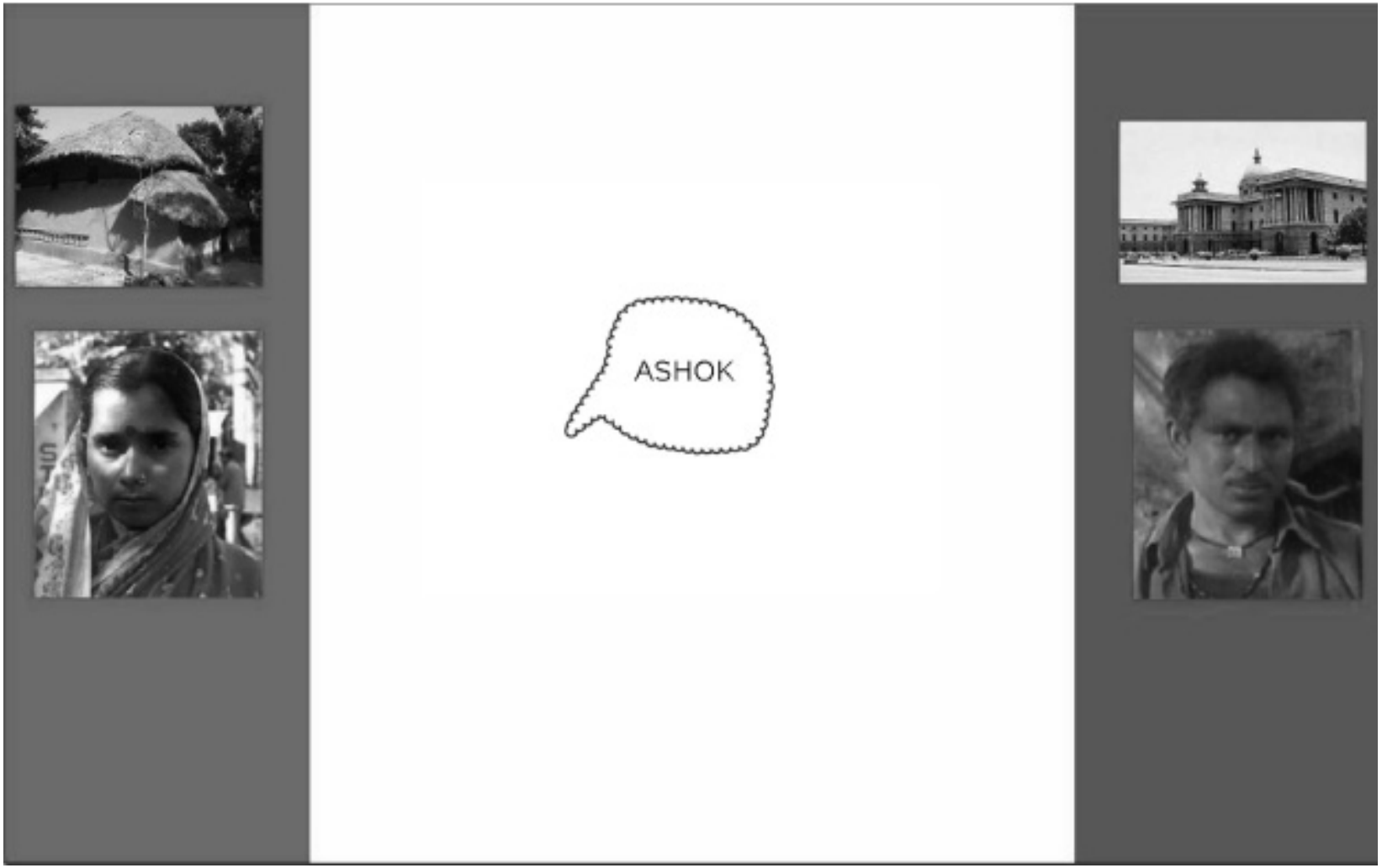
Women's Empowerment?



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?