Theory of Change & Measurement
Course overview

1. Why Evaluate
2. Theory of Change & Measurement
3. Why & When to Randomize
4. How to Randomize
5. Sample Size & Power
6. Ethical Considerations for Randomized Evaluations
7. Threats & Analysis
8. Randomized Evaluation from Start to Finish
9. Applying & Using Evidence
10. The Generalizability Framework
Learning objectives

• Learn how theories of change can serve as a roadmap for program evaluation and a foundation for measurement strategy.

• Discuss fundamental concepts of measurement and considerations when choosing indicators and data sources.

• Understand advantages and challenges of using administrative data.
Outline

I. Theory of Change
II. Measurement concepts
III. Sources of data
IV. Data collection
Program evaluation builds on good theory, program design, and implementation

Strong **theory** and **policy goals** guide identification of needs and logical steps of the ToC

1. Conduct a needs assessment
2. **Design program and build a Theory of Change**
3. Implement and conduct a process evaluation
4. Estimate impact
5. Conduct a cost-effectiveness analysis
Program evaluation builds on good theory, program design, and implementation

THEORY OF CHANGE

Decide on a program to address the identified needs

• What are the inputs or activities?

What steps are needed for the program to achieve the desired change in outcomes?

• What assumptions need to hold?
A Theory of Change (ToC) identifies the causal link between the intervention and the final outcome.
Needs Assessment
• What is the core problem/need?
• What are the contributing factors?
Theory of Change

Intervention/Input

- What are the inputs of the program?
- Example: Textbooks are given to schools
Theory of Change

Needs Assessment

Intervention/Input

Output 1

Output 2

Intermediate outcome 1

Intermediate outcome 2

Intermediate outcome 3

Final outcome 1

Final outcome 2

Output

- What is needed in order for the intervention to be delivered as intended?
- Example: Students receive textbooks from schools
Theory of Change

Intermediate outcomes

- What needs to happen for the final outcome to be affected? (typically a behavior change)

- Example: Increase in students who have passing test scores for the semester
Theory of Change

Final outcomes

- What do you ultimately want to affect with the intervention?
- Example: Increase in high school graduation rates
Theory of Change

Needs Assessment

Intervention/Input

Output 1

Output 2

Intermediate outcome 1

Intermediate outcome 2

Intermediate outcome 3

Final outcome 1

Final outcome 2

Underlying Assumptions

J-PAL | THEORY OF CHANGE AND MEASUREMENT
The intervention is successfully implemented as planned and produces the expected outputs. The intervention’s outputs prompt the expected change in behavior, belief, or knowledge. The change in behavior, belief, or knowledge creates the desired impact.

The needs we are working to solve actually exists, and our understanding of the causes is accurate. The inputs we provide are enough to successfully implement our intervention.
Why spend time on a Theory of Change?

1. **Helps design the intervention**
   - Can be done by thinking of final outcomes first and working backwards
   - Is each step really credible?

2. **Helps design the evaluation by:**
   - Generating research questions
   - Deciding which outcomes to collect and data to measure the outcomes
     - By measuring the right intermediate outcomes, we know more about how our program works
     - Allows to understand the “why,” thus giving richer policy lessons
     - Gives more generalizable knowledge
Theory of Change

- Needs Assessment
  - Intervention/Input
    - Output 1
      - Intermediate outcome 1
      - Final outcome 1
    - Output 2
      - Intermediate outcome 2
      - Final outcome 2
      - Intermediate outcome 3
  - Process Evaluation
    - Assess if the program was implemented as designed
  - Impact Evaluation
    - Assess if the program achieved goals
Example: Legal Aid in Child Welfare

Legal aid program for children who are in foster care, with a priority toward children who are in group homes.

**Background:**

Child protective services are common

- 37% of children in the US are investigated; 5% of children placed in foster care

Children are vulnerable

- e.g. 7x higher rate of mental illness; 3x higher mortality

Policy concerns over bureaucratic frictions:

- Negative correlation between length of stay and child wellbeing.
Example: Legal Aid in Child Welfare

Legal aid program for children who are in foster care, with a priority toward children who are in group homes.

Setting:
Chile – children who live in institutions face a long process before they are reunited with their families or adopted.

Program:
“Mi Abogado” (My Lawyer) provides children with three supports:

- A lawyer (with lower caseload),
- A psychologist, and
- A social worker, with the goals of protecting the rights of children, promoting their return to family life (whether of origin, extended family or through an adoption process).

Children’s cases are assigned to the program by a family-court judge, and the program has a monitoring process of up to 90 days once a child leaves the foster care system.
Children in foster care have a long process before being placed with families.

Children are assigned to a lawyer, psychologist, and a social worker from “Mi Abogado.”

Improvement in legal representation and case management.

Children have a speedy exit from foster care into a more permanent family.

Child/family access/adherence to services.

Intermediate Outcomes

- Increased child safety
- Declined criminal justice involvement
- Increased school attendance

Final Outcomes

Needs

Inputs

Outputs
Why spend time on a Theory of Change?

• Forces us to think critically about the “why,” thus giving richer policy lessons and more generalizable knowledge

• Helps interpret the results of the evaluation
  – If no effect: At what step did the hypothesized causal chain break?
  – If positive effect: What (do we believe) is the mechanism through which the program works?

• Helps design the program and evaluation by:
  – Generating research questions
  – Deciding which data to collect
Discussion question:

What happens if we find no impact?
Discussion question:

What happens if we find no impact?

What happens if we do find an impact?
Without a Theory of Change: Unknowns

If impact evaluation reveals positive impact:

You have learned that program changed specific outcome in specific context under specific conditions.

How do you know if effects will generalize?

If impact evaluation reveals no impact:

You have learned that program did not change specific outcome in specific context under specific conditions.

How do you know how to improve program?

Children are assigned to a lawyer, psychologist, and a social worker from “Mi Abogado”

Final outcomes

Increased child safety
Declined criminal justice involvement
Increased school attendance

Data
You need to know what happens at each stage

Inputs
- Children are assigned to a lawyer, psychologist and a social worker from “Mi Abogado”

Outputs
- Improvement in legal representation and case management
- Children have a speedy exit from foster care into a more permanent family
- Child/family access/adherence to services

Intermediate Outcomes
- ?

Final outcomes
- Increased child safety
- Declined criminal justice involvement
- Increased school attendance
You need **data** on what happens at each stage

**Inputs**
- Children are assigned to a lawyer, psychologist and a social worker from “Mi Abogado”

**Outputs**
- Improvement in legal representation and case management

**Intermediate Outcomes**
- Children have a speedy exit from foster care into a more permanent family
- Child/family access/adherence to services

**Final outcomes**
- Increased child safety
- Declined criminal justice involvement
- Increased school attendance
Outline

I. Theory of Change

II. Measurement concepts

III. Sources of data

IV. Data collection
Concepts of measurement

Construct

The main concept being investigated. A construct is often abstract. (E.g., child safety).

Indicators

How you actually measure or “operationalize” your construct. (E.g., reports of child abuse and/or neglect).

Data collection

Data

What we use to measure our indicators.
Child wellbeing is:

A. A construct
B. An indicator
C. Data
D. Don’t know
Criminal justice involvement is:

A. A construct
B. An indicator
C. Data
D. Don’t know
Arrest records are:

A. A construct
B. An indicator
C. Data
D. Don’t know
How do you choose which indicators to measure?

• In many cases, there are several indicators that can be mapped back to the same construct.
• For example, consider the construct of child safety, we could measure:
  – Child return to foster care
  – New investigations opened for child maltreatment
  – Quarterly observations on whether the child is a victim of a crime or children reported as missing

• Two criteria
  – Validity
  – Reliability
Measurement criteria

- **Validity** – measuring the right thing
- **Reliability** – measuring the thing precisely
Validity (a.k.a. accuracy or unbiasedness)

Construct

How well does the indicator map to the construct we are trying to measure?

Indicators

Can we collect data on the indicator in an unbiased way?

Data collection

Data

Theory

LOW

HIGH
Question:

Construct: Students’ school performance
Indicator: School attendance rates

Where does that indicator land?
Reliability (a.k.a. precision)

**Construct**
- How time/situation specific is the indicator?

**Indicators**
- Is the indicator measured in a way that is consistent and precise? Would we get the same data if we measured several times?

**Data collection**

**Data**

Theory
Question:

Construct: Students’ school performance
Indicator: School attendance rates

Where does that indicator land?
The goals of measurement

![Diagram showing the goals of measurement with axes for Reliability and Validity, where high and low values are depicted.](image)
Maximizing validity

**Theory:** Think about how the indicator maps to the construct

**Practice:** Make sure data is collected in a way that is not systematically biased

---

**Possible sources of bias:**
- Theory/mapping
- Experimenter demand effects
- Social desirability bias
- Recall bias
- Translation/interpretation

**Possible ways to minimize bias:**
- Use administrative data where possible
- Use methods for collecting data on sensitive topics
- Back translation and piloting
- Use multiple data sources
Maximizing reliability

**Theory:** Think about how time/situation-specific the indicator is

**Practice:** Make sure to use indicators that have been validated (as much as possible)

### Possible sources of unreliability:
- Fatigue
- Ambiguous wording (e.g. “# people in household”)

### Possible ways to maximize reliability:
- Consider survey length
- Consider the answer choices
- [Piloting](#)
- [Training of survey staff](#)
- [Conduct data quality checks](#)

---

**Reliability** (a.k.a. precision)
Question:
Which is the worst?

A. Low validity, low reliability
B. Low validity, high reliability
C. High validity, low reliability
D. All equally bad
E. Don’t know/can’t say
Outline

I. Theory of Change
II. Measurement concepts
III. Sources of data
IV. Data collection
Where can we get data?

• From existing sources *(Secondary data)*
  – Publicly available datasets
  – Sensor data
  – Web traffic data
  – Existing survey data
  – Administrative data

• Collected by researchers *(Primary data)*
  – Surveys
  – Focus groups
  – Observation
  – Games
  – Diaries
# Types and sources of data

<table>
<thead>
<tr>
<th>Information about a person/household</th>
<th>NOT about a person/household</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information reported by a person</td>
<td></td>
</tr>
<tr>
<td>- Cognition, anthropometrics</td>
<td>- Farming inputs and outputs</td>
</tr>
<tr>
<td>- Demographics</td>
<td>- Quality of medical care</td>
</tr>
<tr>
<td>- Behavior, beliefs</td>
<td>- Business income taxes</td>
</tr>
<tr>
<td>- Patience, risk aversion, psychometrics</td>
<td></td>
</tr>
<tr>
<td>- Knowledge</td>
<td></td>
</tr>
<tr>
<td>- Income, expenditure</td>
<td></td>
</tr>
<tr>
<td>Automatically generated</td>
<td></td>
</tr>
<tr>
<td>- Bank transactions</td>
<td>- Prices</td>
</tr>
<tr>
<td>- Phone data</td>
<td>- Weather, air quality</td>
</tr>
<tr>
<td>- Sales records</td>
<td>- Stock markets</td>
</tr>
<tr>
<td>- School/university records, criminal record</td>
<td>- VAT records</td>
</tr>
</tbody>
</table>
Information collected, used, and stored primarily for administrative (i.e., operational) purposes, rather than research purposes.

- Medical records
- Grade books
- Arrest records
- Bank account data
- Personnel records
- Log books
The promise of administrative data in RCTs

Data quality
- Objective
- Representative
- Robust to attrition

Sample sizes
- More treatment arms possible
- More precise estimates

Higher speed and lower cost
- Provides a ready sampling frame for surveys, replacing need for a full census/listing
- May replace costly field data collection

Innovative uses
- Descriptive statistics for pilots and policy analysis
- Hybrid studies combining administrative with survey data
- Completely new data and unique measurement tools
- Machine learning / AI tools overlaid on admin data
Question:
Which data sources are most relevant for the programs/policies you work on?

A. Administrative data
B. Survey data
C. Both
D. Other
Outline

I. Theory of Change
II. Measurement concepts
III. Sources of data
IV. Data collection
Concepts of measurement

- **Construct**
- **Indicators**
- **Data**

Theory → **Construct** → **Indicators** → **Data**

Data collection

Decisions about collection method (e.g., interview or observation)
The Response Process

1. Comprehension of the question
2. Retrieval of Information
3. Judgement and Estimation
4. Reporting an Answer
Step 1: Comprehension

How many fruits and vegetables did your child eat yesterday?
Breakfast: Apple, orange juice, oatmeal
Lunch: Cheese, sweet potato, crackers, crackers
Dinner: ?
Step 3: Estimation/Judgement

1. Comprehension of the question
2. Retrieval of Information
3. Judgement and Estimation
4. Reporting an Answer

2 fruits, 0 vegetables at breakfast, 0 fruits, 1 vegetable at lunch... Maybe 2 vegetables at dinner?
How many fruits and vegetables did your child eat yesterday?
Summary: Potential sources of measurement error

Question issues
- Arise due to poorly worded questions
- For example: vagueness, presumptions, and framing effects

Response-related issues
- Caused by using incomplete or overlapping categories in responses
- For example: errors in how answer choices are constructed

Respondent issues
- Arise due to the respondent’s inherent bias while answering questions
- For example: recall bias, social desirability bias, or reporting bias
Example: Legal Aid in Child Welfare

Legal aid program for children who are in foster care, with a priority toward children who are in group homes.

Outcomes & measurement

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Measurement</th>
<th>Data source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child safety</td>
<td>- Child return to foster care&lt;br&gt;- New investigations opened for child maltreatment&lt;br&gt;- Quarterly observations on whether the child is a victim of a crime, or children reported as missing</td>
<td>Admin data - SENAME</td>
</tr>
<tr>
<td>Criminal justice involvement</td>
<td>Number of crimes reported where a child from the program is suspected of committing a crime, each quarter</td>
<td>Admin data - Judiciary Registry</td>
</tr>
<tr>
<td>School attendance</td>
<td>Share of school days that the child attended school</td>
<td>Admin data - Ministry of Education</td>
</tr>
</tbody>
</table>
Recap and concluding thoughts

• A Theory of Change is a useful initial step for any type of program evaluation.

• A Theory of Change helps inform which inputs, outputs, and intermediate outcomes are needed to understand how/why a program did (or did not) work.

• For best results, all steps of the Theory of Change need to be measured, and measurement needs to be done carefully.

• The process of collecting “good” data requires a lot of effort and thought and involves tradeoffs.
  – Quality vs cost
  – Validity vs reliability
References

Resources for further reading

J-PAL Research Resource: [Introduction to measurement and indicators](#)

J-PAL Research Resource: [Repository of measurement and survey design resources](#)
Reuse and citation

To reference this lecture, please cite as:


This lecture is made available under a Creative Commons Attribution 4.0 License (international): https://creativecommons.org/licenses/by/4.0/