J-PAL
ABDUL LATIF JAMEEL POVERTY ACTION LAB

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. Randomized Evaluation from Start to Finish
7. Threats & Analysis
8. Ethical Considerations
9. Generalizing & Applying Evidence

Learning objectives

- Learn how theories of change can serve as a roadmap for program evaluation and a foundation for measurement strategy
- Understand how a theory of change can guide your measurement strategy and help you interpret the results of your evaluation
- Discuss fundamental concepts of measurement and considerations when choosing indicators and data sources
- Understand different types of data and their advantages and challenges

Outline

I. Theory of Change

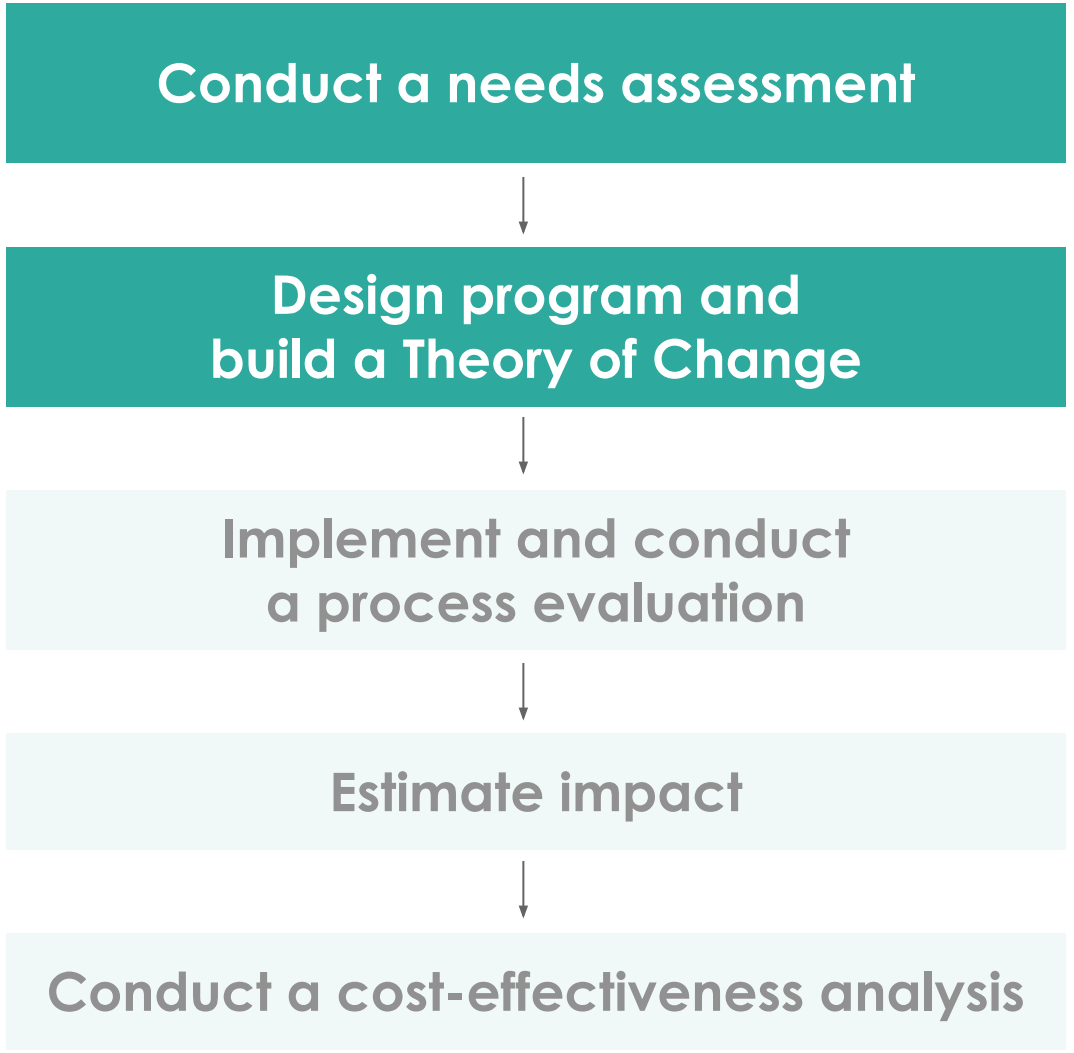
II. Measurement concepts

III. Sources of data

IV. Data collection

Evaluation builds on thoughtful and informed program design

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



A good impact evaluation builds on good program design that identifies and addresses a local need

NEEDS ASSESSMENT

What is the problem?

- Understand what is the extent of the problem
- Identify the barriers that exist and levers that are promising around the problem
- Identify who is the most in need

What are its contributing factors?

- Build a hypothesis about what causes the problem and propose possible solutions

Data driven
(E.g., administrative data, qualitative information)

A good evaluation builds on a well-developed Theory of Change

THEORY OF CHANGE

Decide on a program to address the identified needs

- What are the inputs or activities?

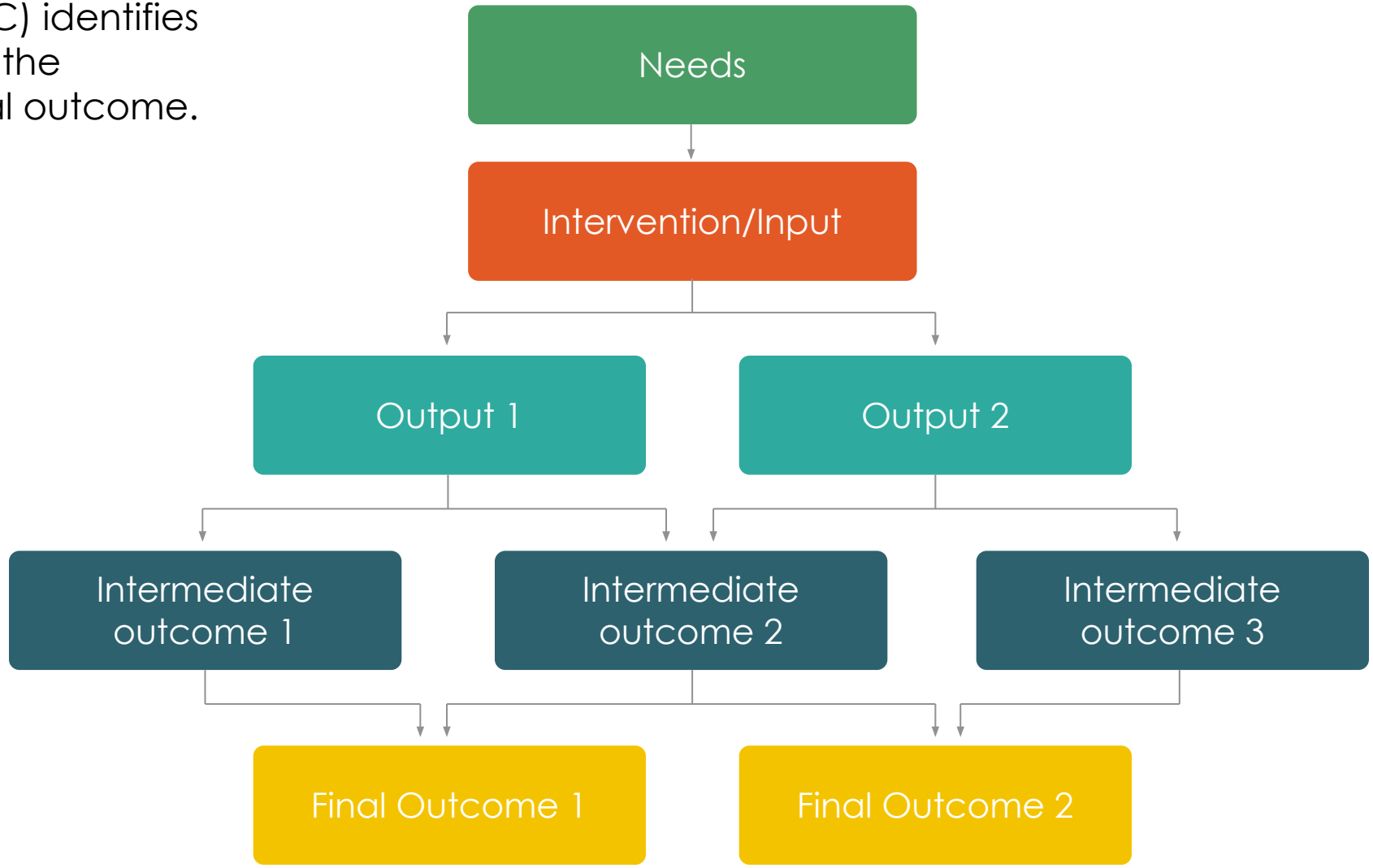
What steps are needed for the program to achieve the desired change in outcomes?

- What assumptions need to hold?

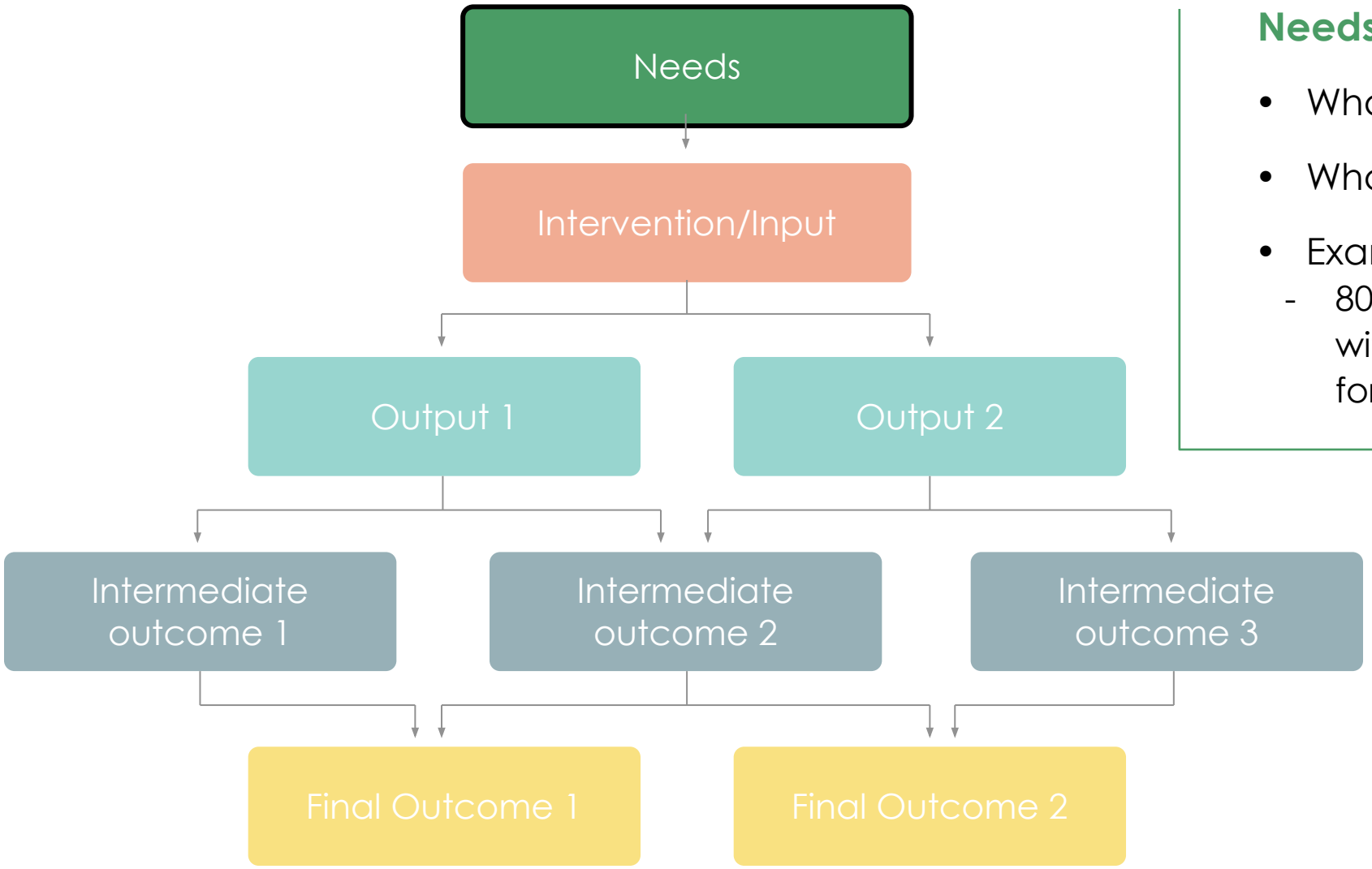
Theory of Change

ESP WORKBOOK

A Theory of Change (ToC) identifies the causal link between the intervention and the final outcome.



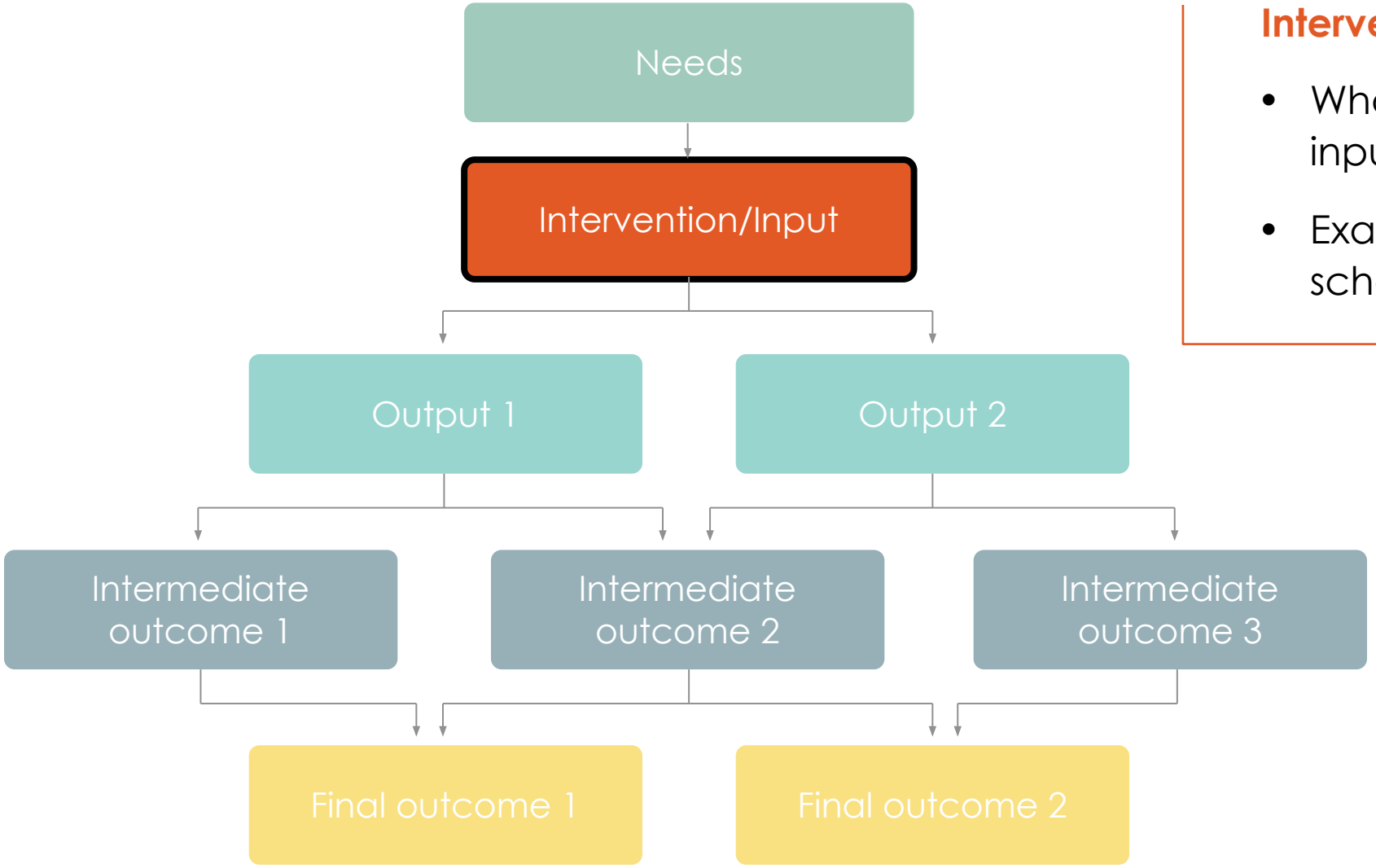
Theory of Change



Needs

- What is the core problem?
- What are the contributing factors?
- Example: Low test scores
  - 80% of students are in classrooms with less than one English textbook for every 20 students.

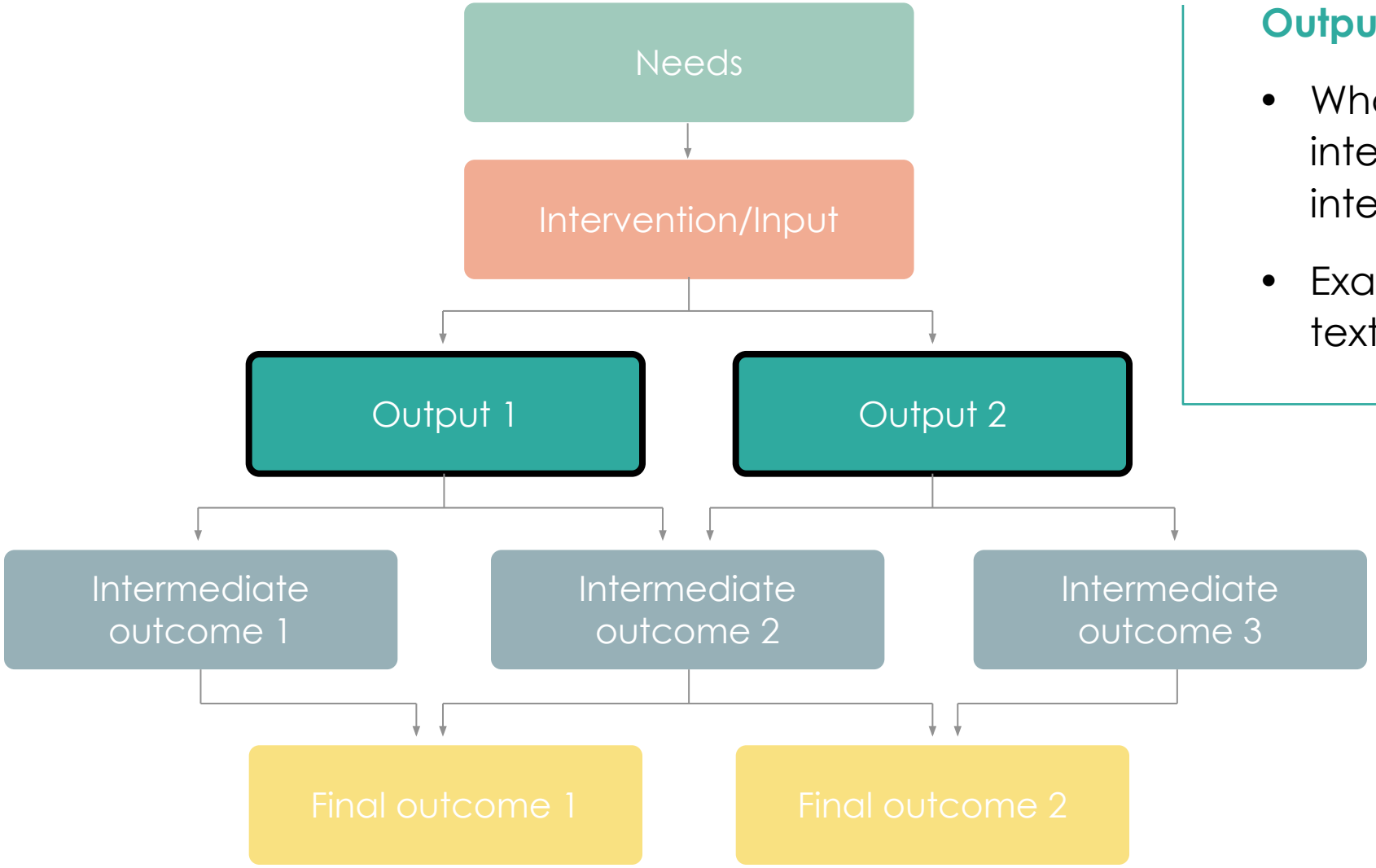
Theory of Change



Intervention/Input

- What are the components or inputs of the program?
- Example: Textbooks are given to schools

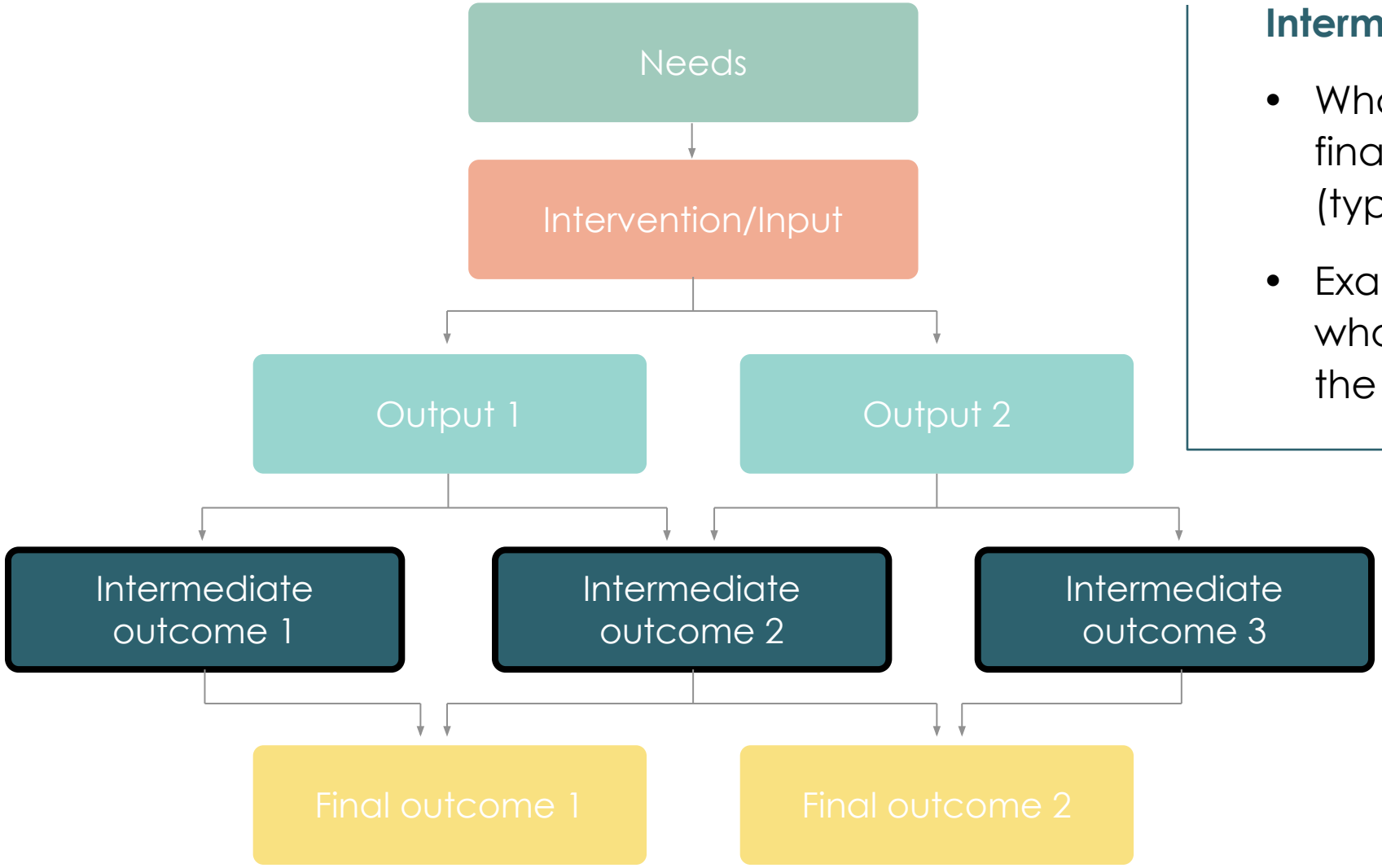
Theory of Change



Output

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

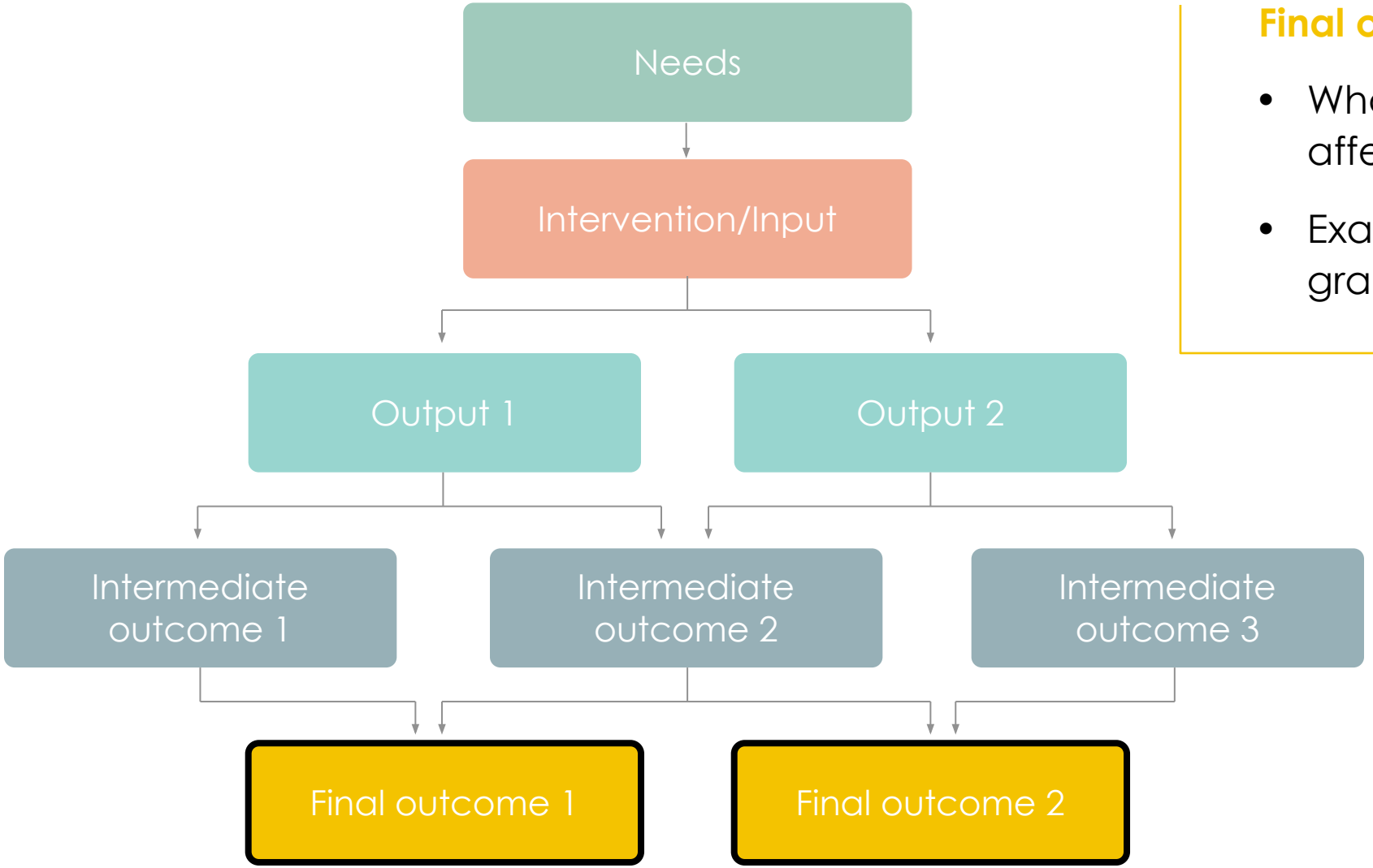
Theory of Change



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

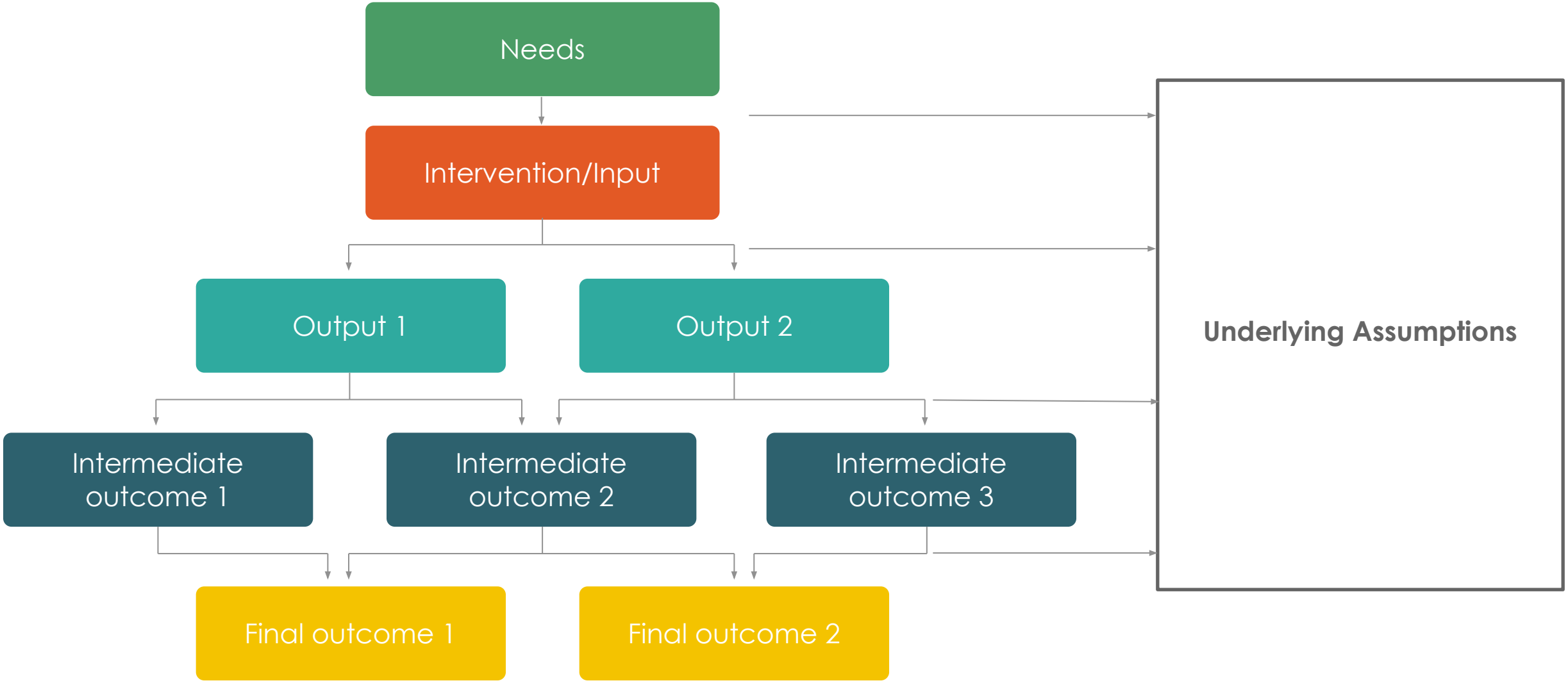
Theory of Change



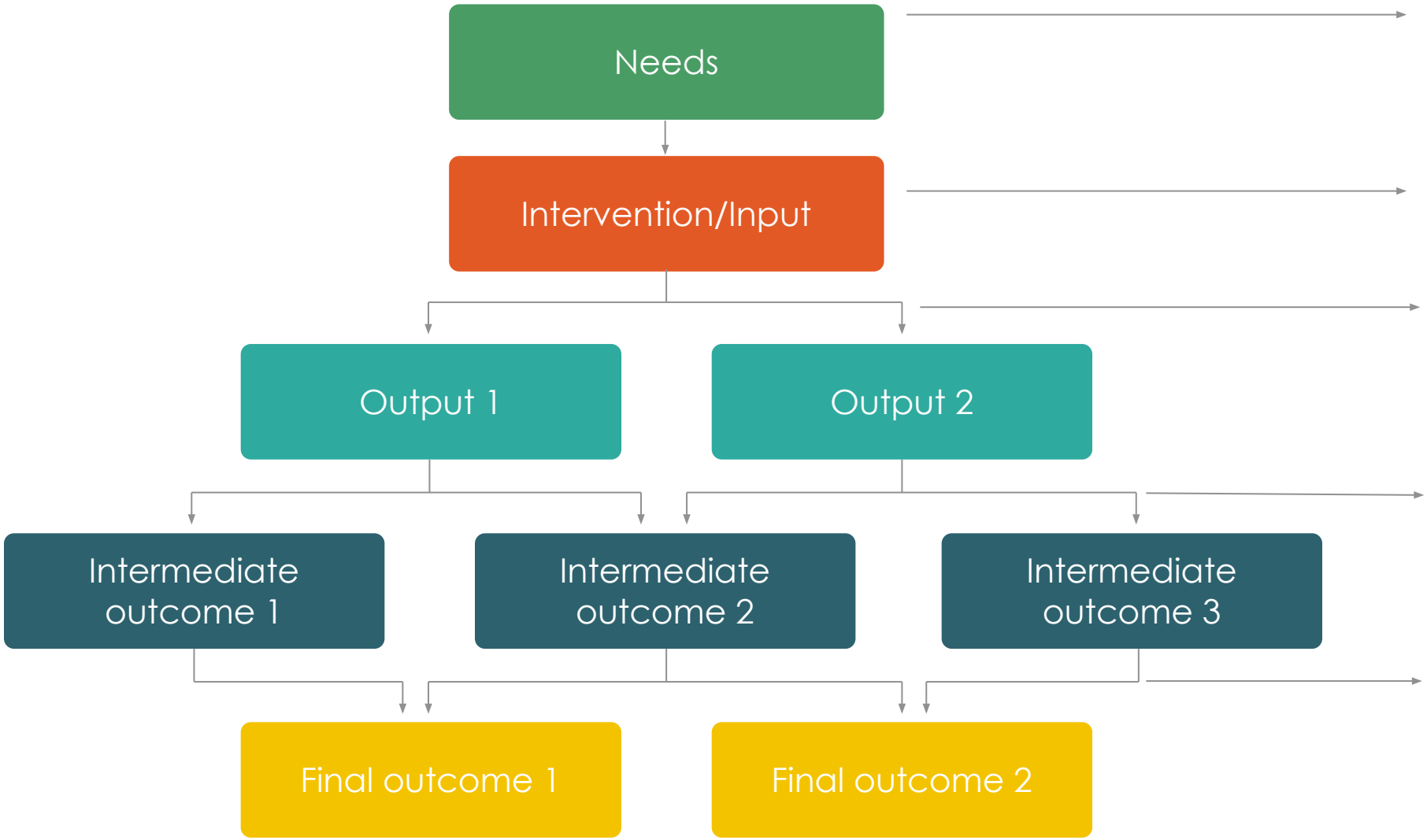
Final outcomes

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

Theory of Change



Theory of Change



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.

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.

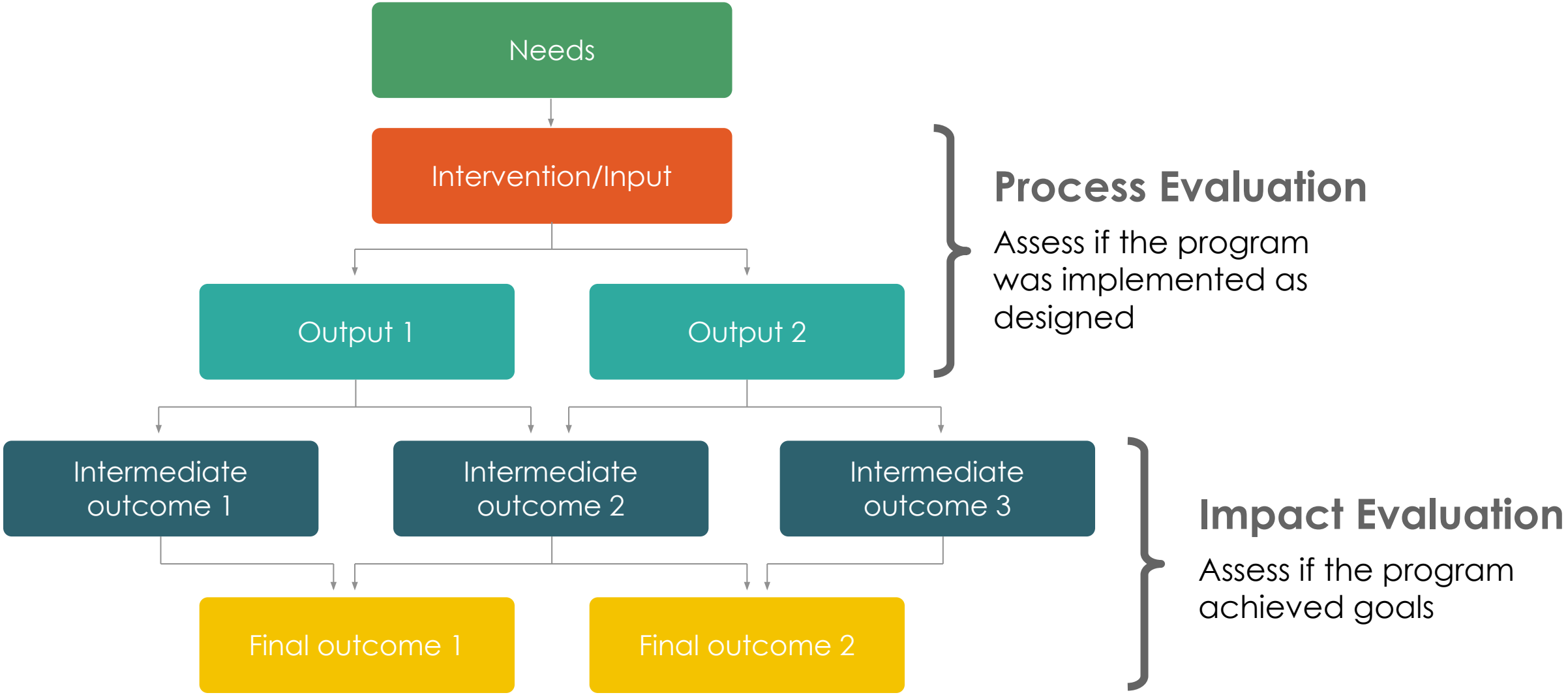
The change in behavior, belief, or knowledge creates the desired **impact.**

Why spend time on a Theory of Change?

1. **Helps design the evaluation and interpret results**
   - Can be done by thinking of final outcomes first and working backwards
   - Is there evidence that makes each step in the TOC credible?

2. **Helps design the evaluation**
   - Generate research questions
   - Decide which outcomes to measure and what data to collect
     - By measuring the right intermediate outcomes, we know more about how our program works
     - Allows us to understand the "why," thus giving richer policy lessons
     - Gives more generalizable knowledge

Theory of Change



Example: Legal aid in child welfare

Background:

Child protective services are common

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

Chile: In 2022, 10,700 children were in substitute care

- Allegations: 84% involved neglect and 28% physical abuse

Children in the foster care system are vulnerable

7x higher rate of mental illness; 3x higher mortality

Chile: In 2019, 41% of children in the system were in institutional homes

Policy concerns over bureaucratic frictions

Negative correlation between length of stay and child wellbeing

Chile: Average stay 3 years, which is higher than other countries

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.

Context of the evaluation:

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

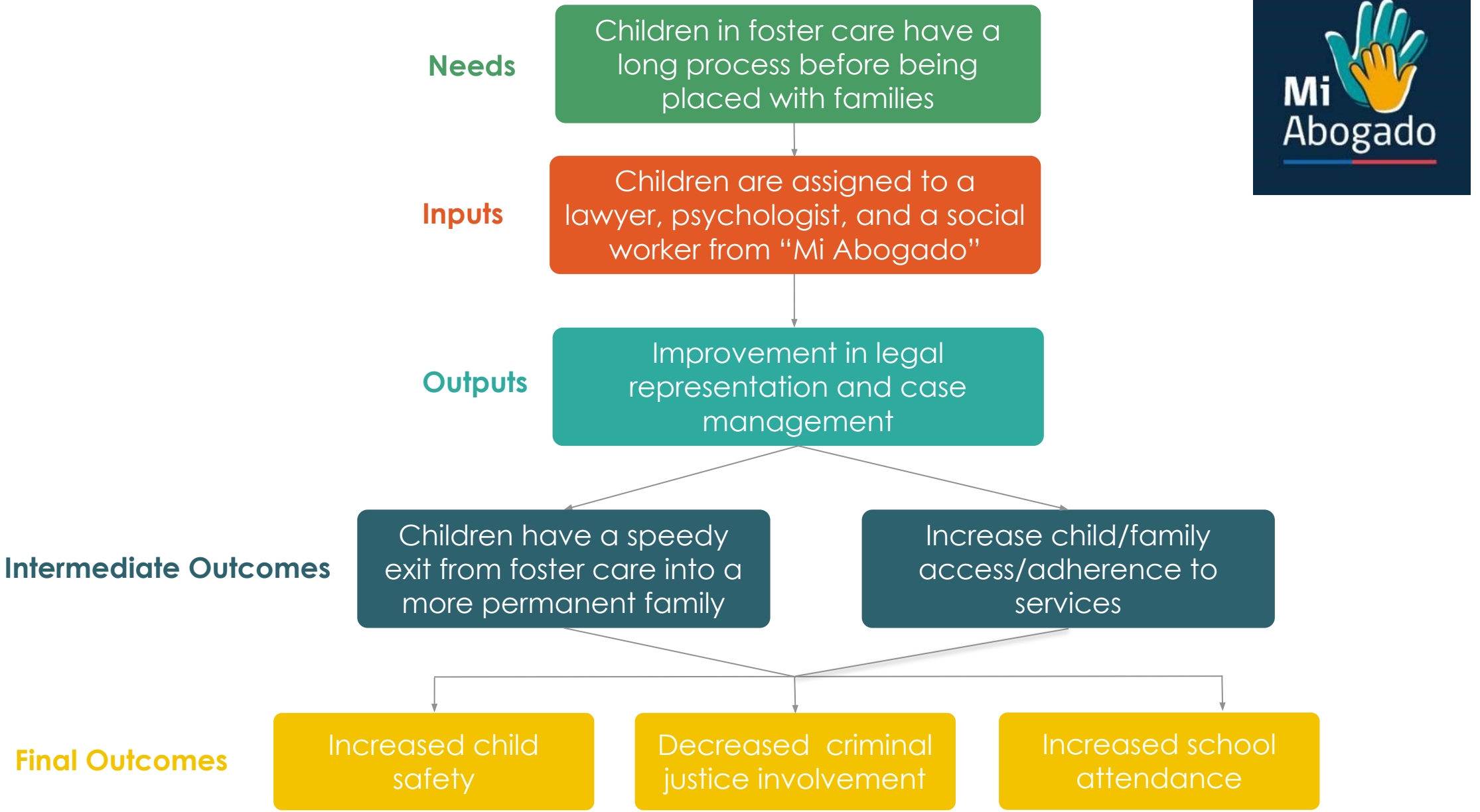
Program:

"Mi Abogado" (My Lawyer) provides children with three supports, with the goals of protecting the rights of children and promoting their return to family life:

- A lawyer with a much lower caseload
- A psychologist
- A social worker

Children's cases are assigned to the program by a family-court judge

Monitoring continues for up to 90 days once a child leaves the foster care system


Needs
Children in foster care have a long process before being placed with families
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
Increase child/family access/adherence to services
Final Outcomes
Increased child safety
Decreased criminal justice involvement
Increased school attendance

A Theory of Change helps to interpret the results of an evaluation

- 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?

Discussion question:

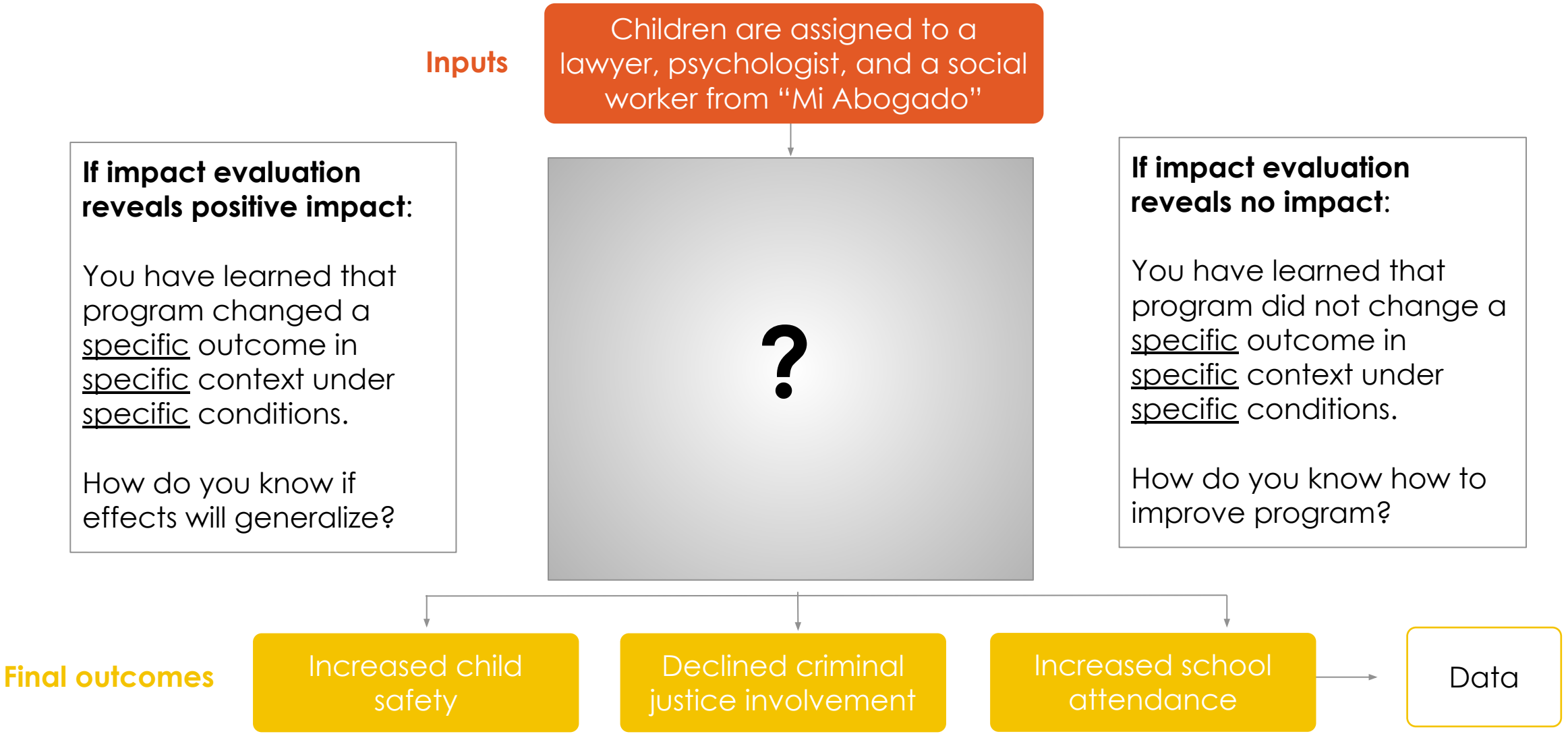
From a policy perspective, what would we do if we find no impact of the program?

Discussion question:

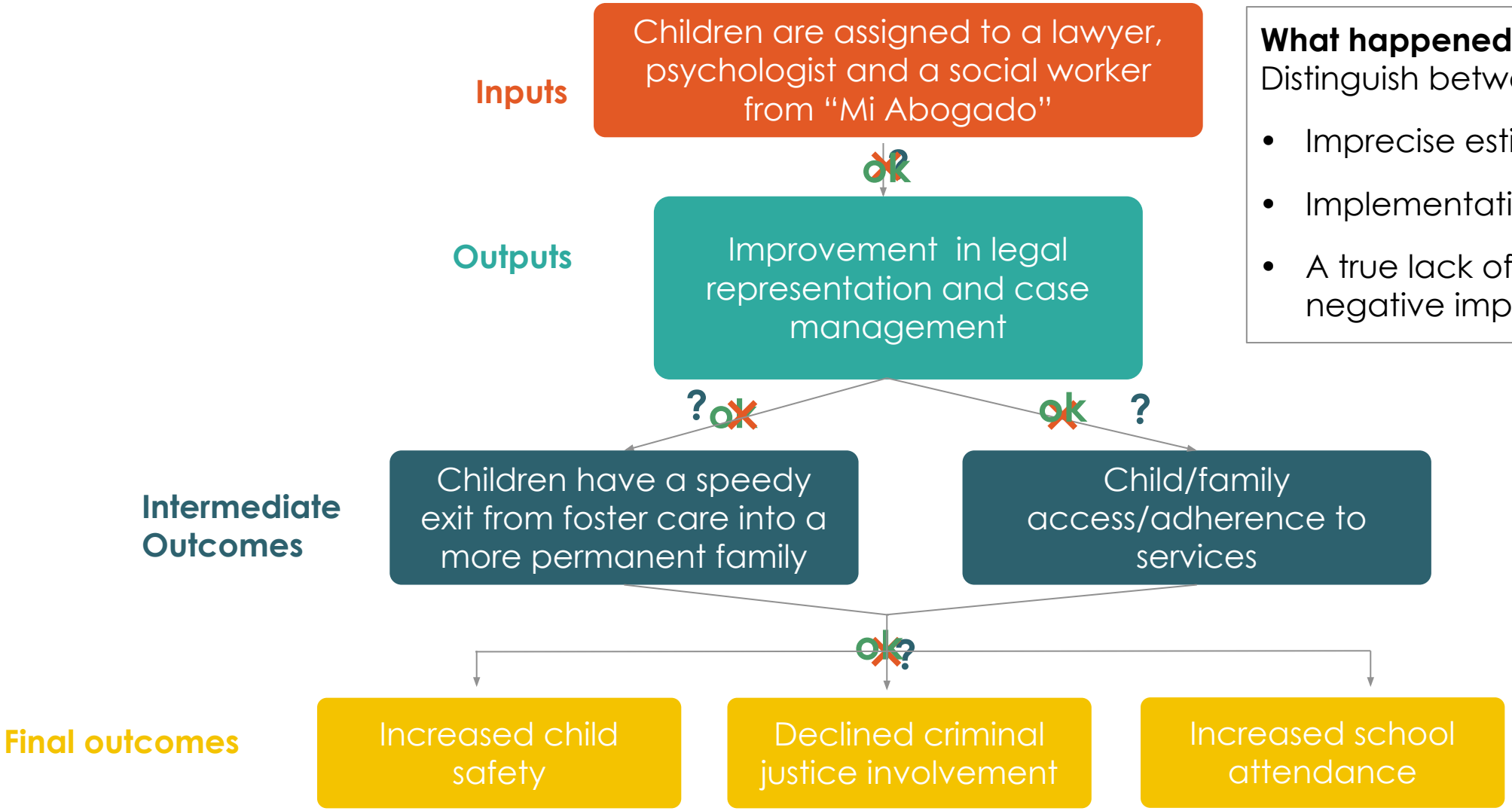
From a policy perspective, what would we do if we find no impact of the program?

What happens if we do find an impact?

Without a Theory of Change: Unknowns



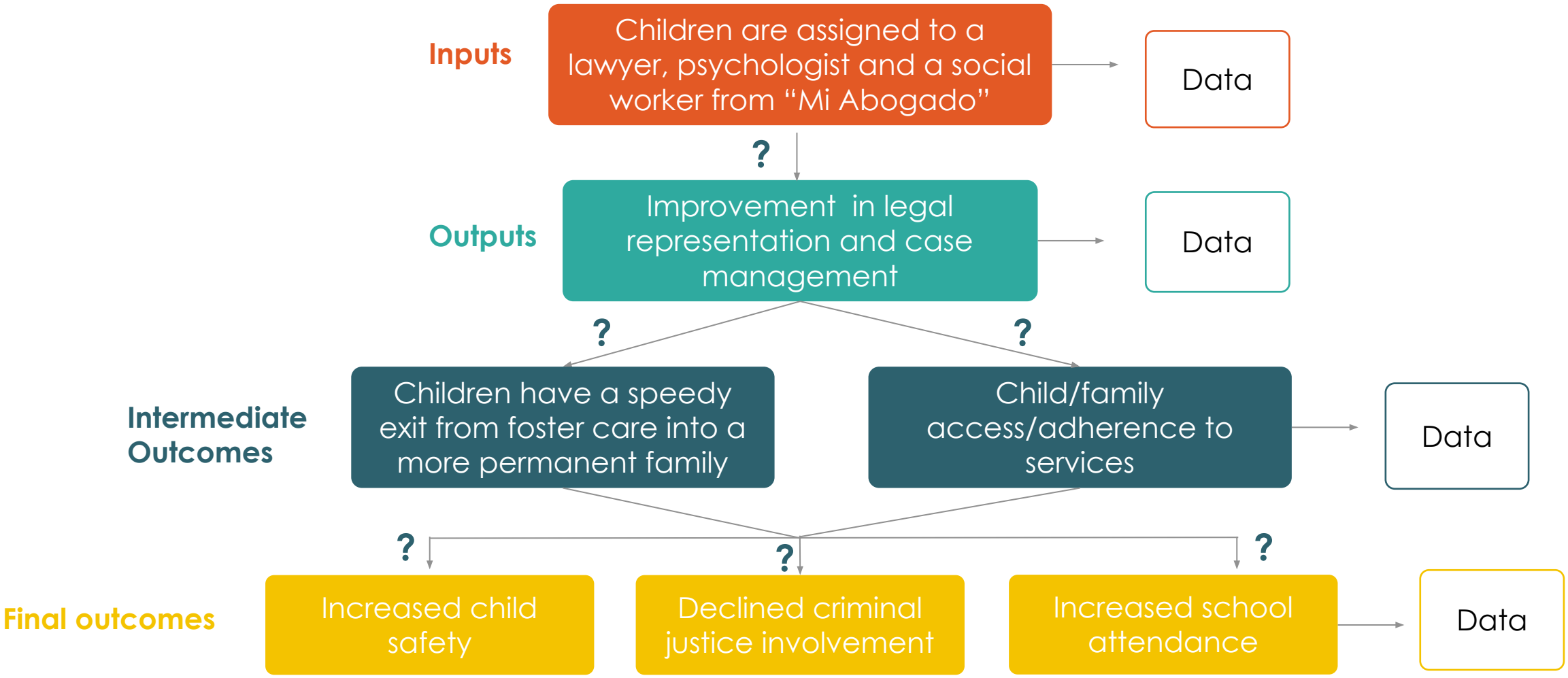
If we find negative or no impact, we can look to the Theory of Change for answers



What happened at each stage?
Distinguish between:

- Imprecise estimates
- Implementation failure
- A true lack of impact or negative impact

You need **data** on what happens at each stage



Example: Camden Coalition's care management program

A high-touch, face-to-face model of health care designed for high-use, high-need patients to connect them to existing services and to manage their care.

Setting

Camden, New Jersey - high-use, high-need patients

Background

Health care costs are heavily concentrated

5% of the population accounts for 50% of spending in a given year

Relevant policy concern

0.5% of Camden's population have frequent hospital admissions, complex medical and social needs

- This population accounts for 11% of the hospital expenditures

Example: Camden Coalition's care management program

A high-touch, face-to-face model of health care designed for high-use, high-need patients to connect them to existing services and to manage their care.

Program's goal

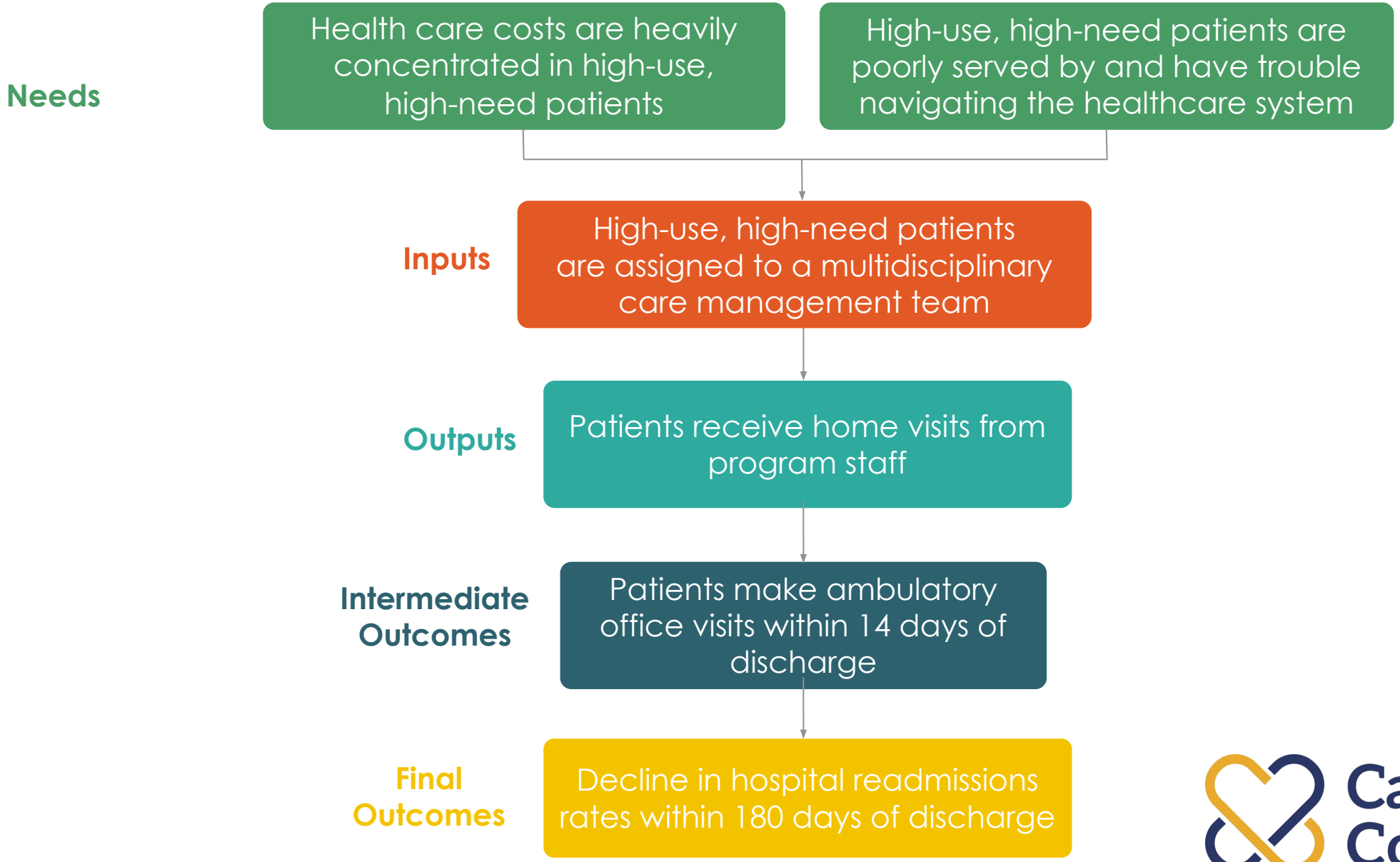
To break the cycle of repeat hospitalizations, improve patients' wellbeing, and reduce health care costs.

Program components

Care management once patients returned home for ~90 days provided by:

- Registered nurses
- Social workers
- Licensed practical nurses
- Community health workers and health coaches

Care management team conducted home visits to help patients with disease and medication management, schedule and accompany them to doctor visits, connect them with social services and other programs.

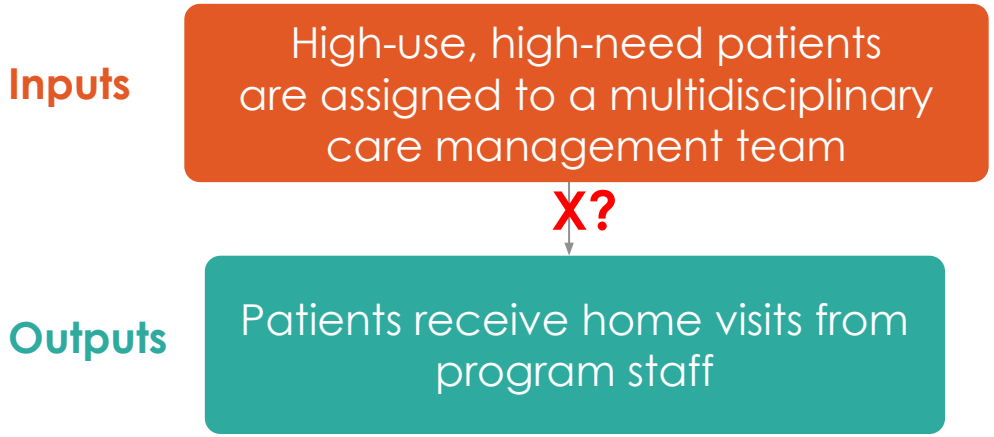


Camden Coalition

Example: Camden Coalition's care management program

- Randomized evaluation of the program found that the program did not reduce hospital readmission!

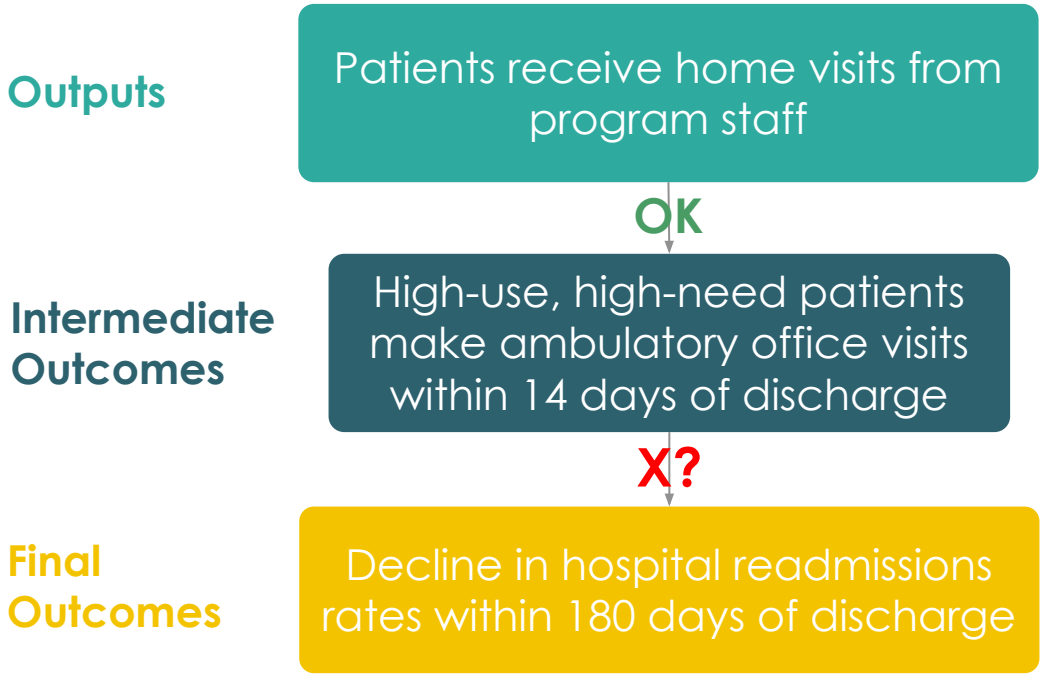
- Two possible explanations
  - Failure of implementation



Example: Camden Coalition's care management program

- Randomized evaluation of the program found that the program did not reduce hospital readmission!

- Two possible explanations:
  - Failure of implementation
  - Program's underlying theory of change was not right



Outline

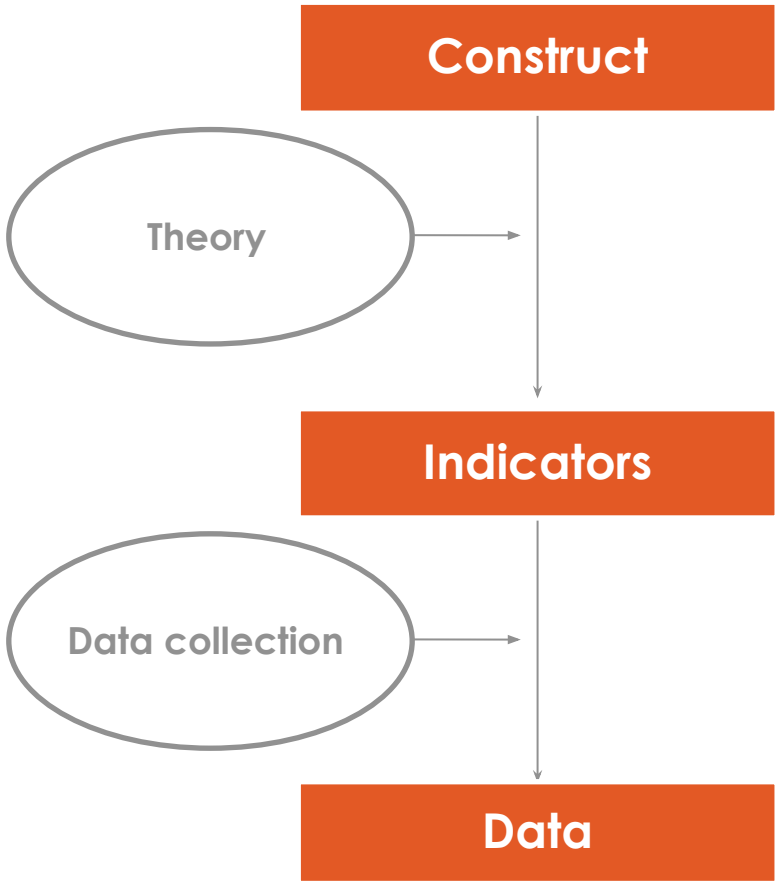
I. Theory of Change

II. Measurement concepts

III. Sources of data

IV. Data collection

Concepts of measurement



The main concept being investigated. A construct is often abstract.

(E.g., child safety).

How you actually measure or "operationalize" your construct.

(E.g., reports of child abuse and/or neglect).

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

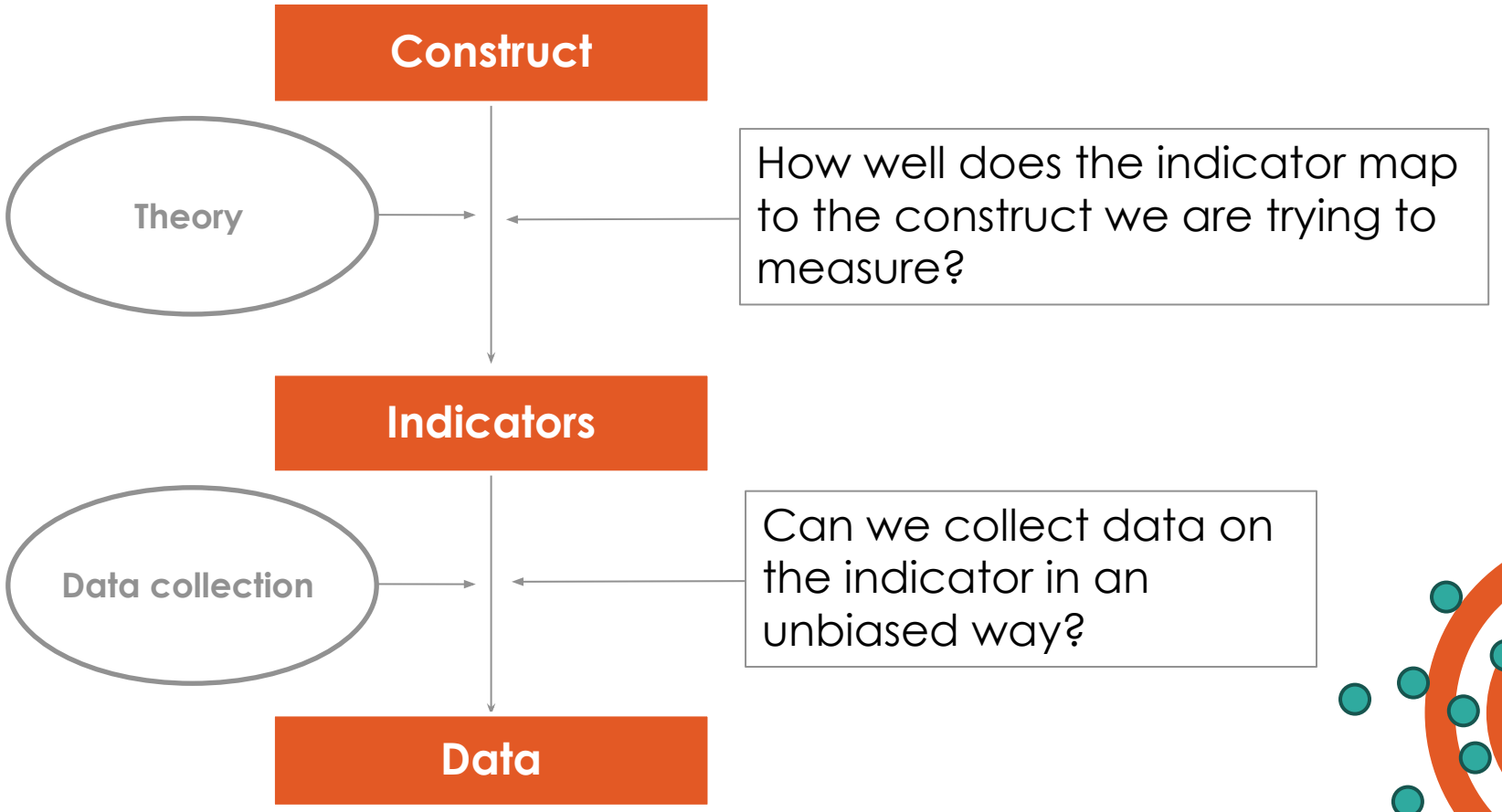
- **Validity** – measuring the right thing

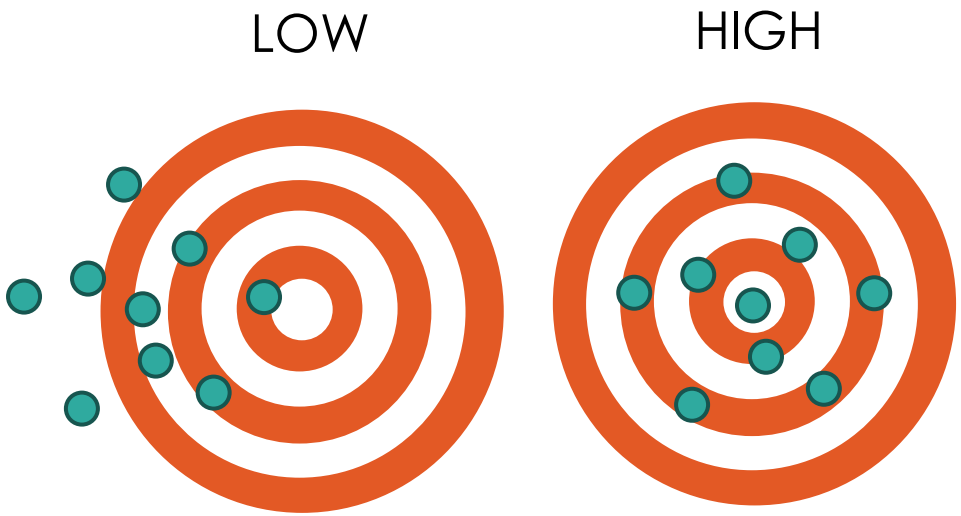
Low  High 

- **Reliability** – measuring the thing precisely

Low  High 

Validity (a.k.a. accuracy or unbiasedness)



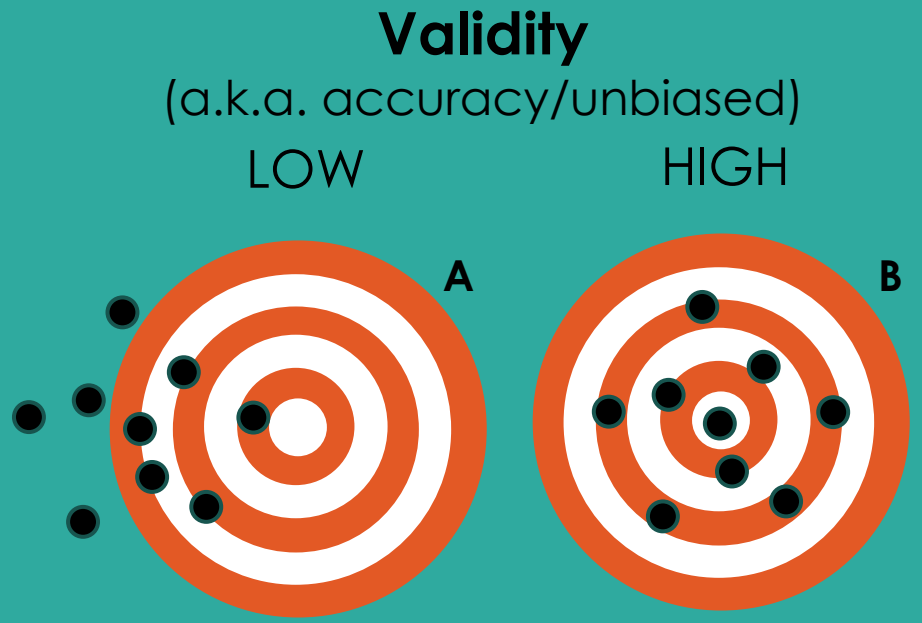


Question:

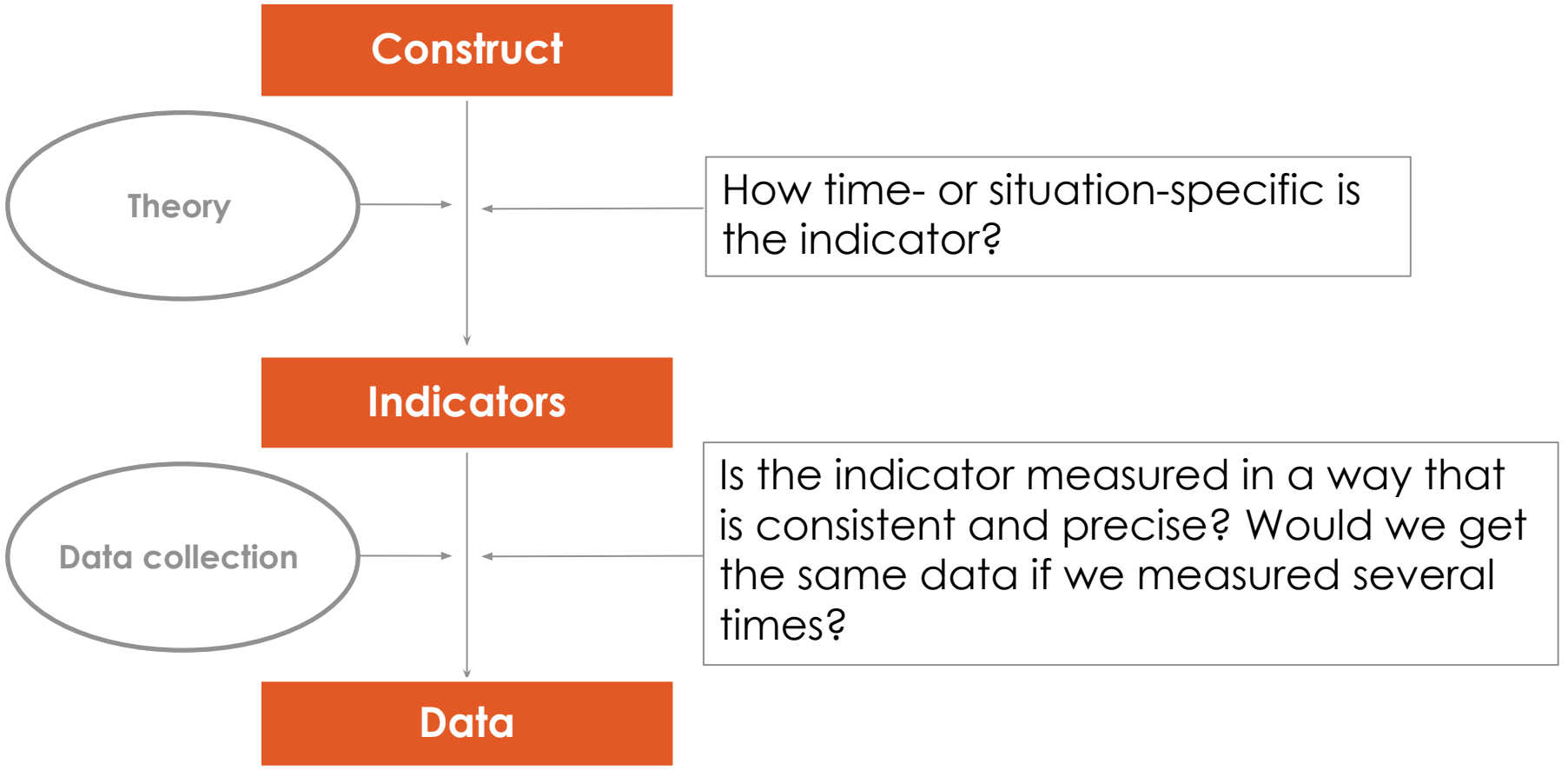
Construct: Students' school performance

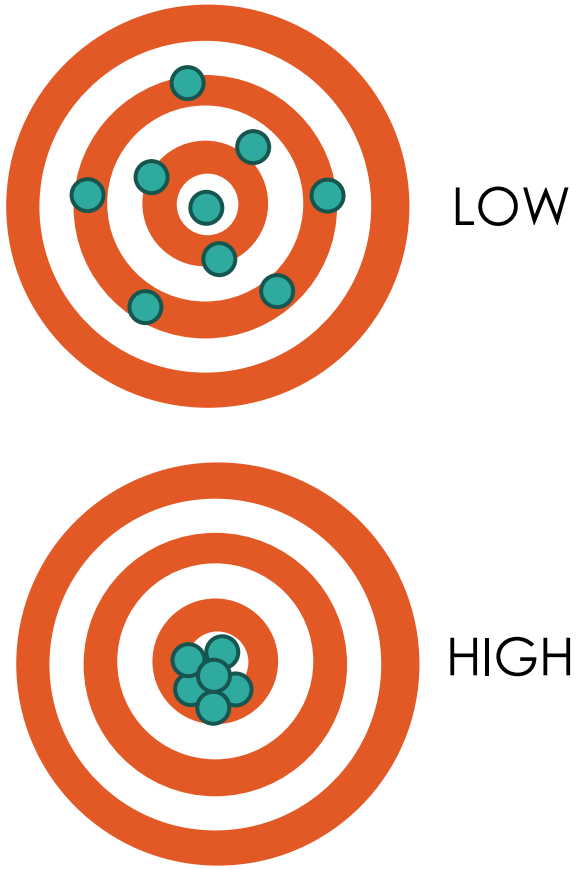
Indicator: School attendance rates

Where does that indicator land?



Reliability (a.k.a. precision)





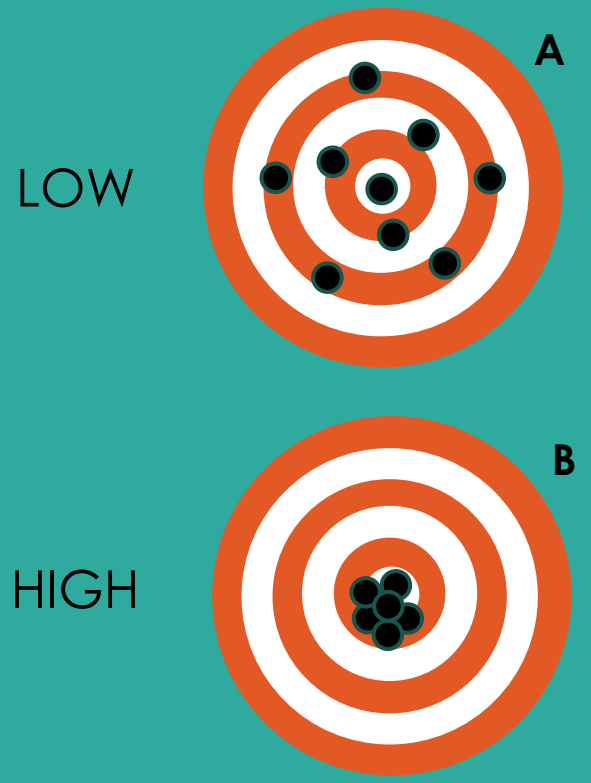
Question:

Construct: Students' school performance

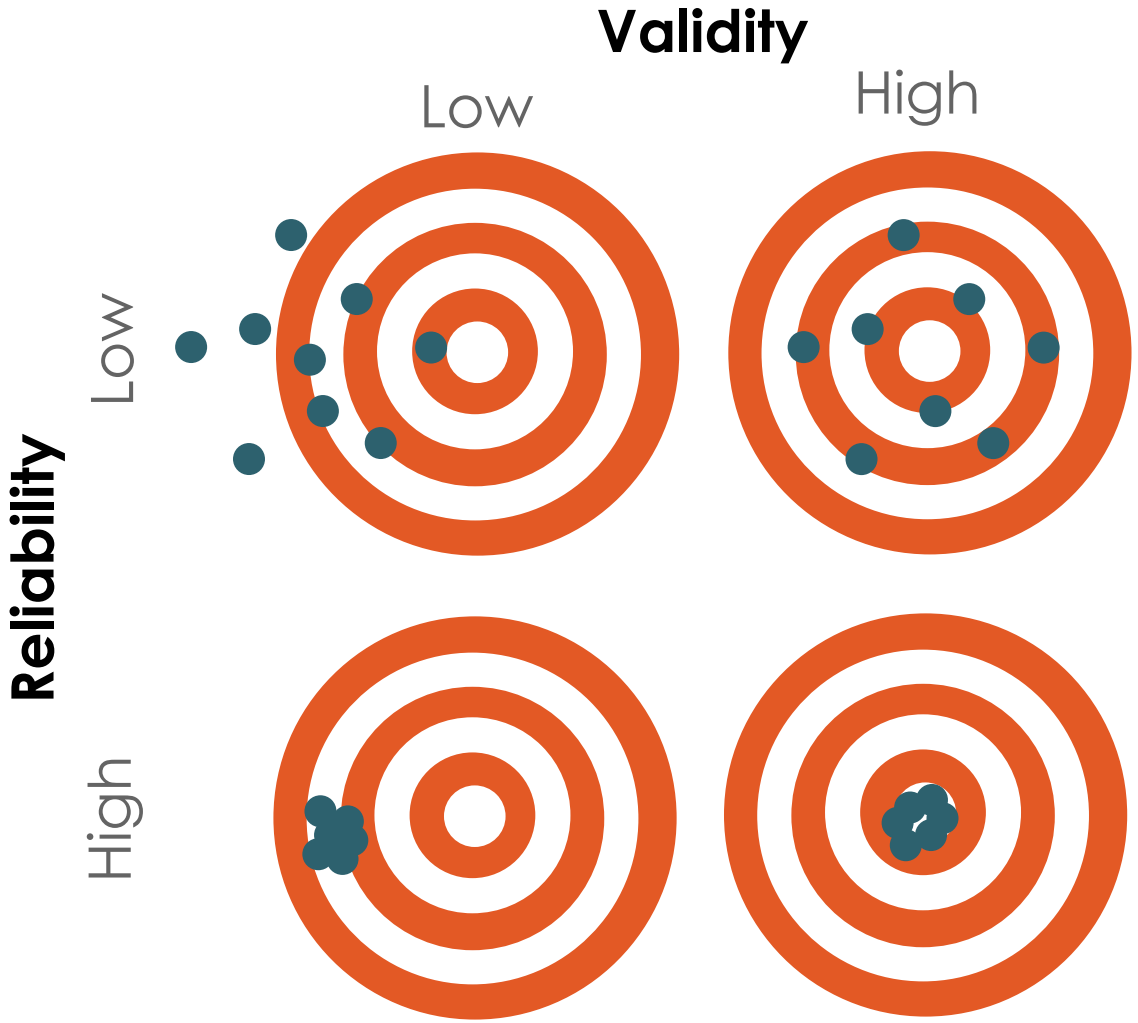
Indicator: School attendance rates

Where does that indicator land?

Reliability
(a.k.a. precision)

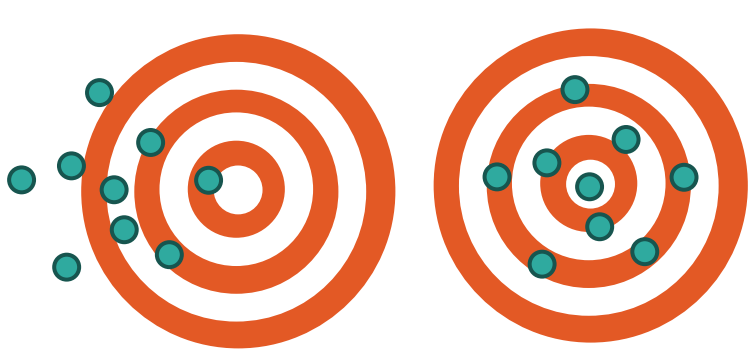


The goals of measurement



Maximizing validity

Validity
(a.k.a. accuracy or unbiasedness)



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- or situation-specific the indicator is

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

Reliability
(a.k.a. precision)

LOW

HIGH

Possible sources of unreliability:	**Possible ways to maximize reliability**
• Fatigue • Ambiguous wording (e.g. "# people in household")	• Consider survey length • Consider the answer choices • Questionnaire piloting • Training of survey staff • Conduct data quality checks

Question:

Which is the worst?

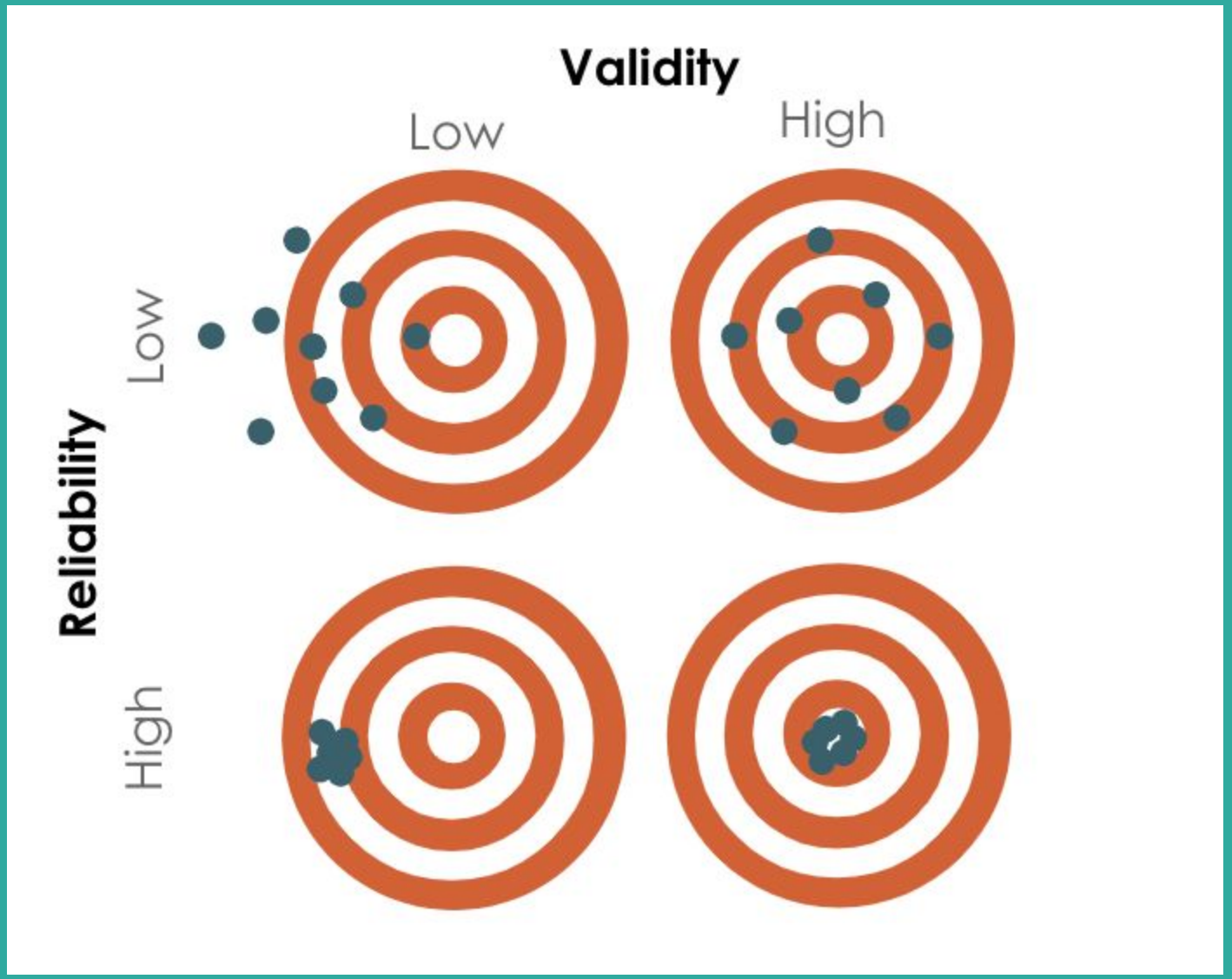
A. Low validity, low reliability

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

	Information about a person/household	Information NOT about a person/household
Information reported by a person	• Cognition, anthropometrics • Demographics • Behavior, beliefs • Patience, risk aversion, psychometrics • Knowledge • Income, expenditure	• Farming inputs and outputs • Quality of medical care • Business income taxes
Information automatically generated	• Bank transactions • Phone data • Sales records • School/university records, criminal record	• Prices • Weather, air quality • Stock markets • VAT records

Administrative data

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. Observational or qualitative data

D. Both

E. Other

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

Outcome	Measurement	Data source
Child safety	- 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	Admin data - SENAME
Criminal justice involvement	Number of crimes reported where a child from the program is suspected of committing a crime, each quarter	Admin data - Judiciary Registry
School attendance	Share of school days that the child attended school	Admin data - Ministry of Education

Outline

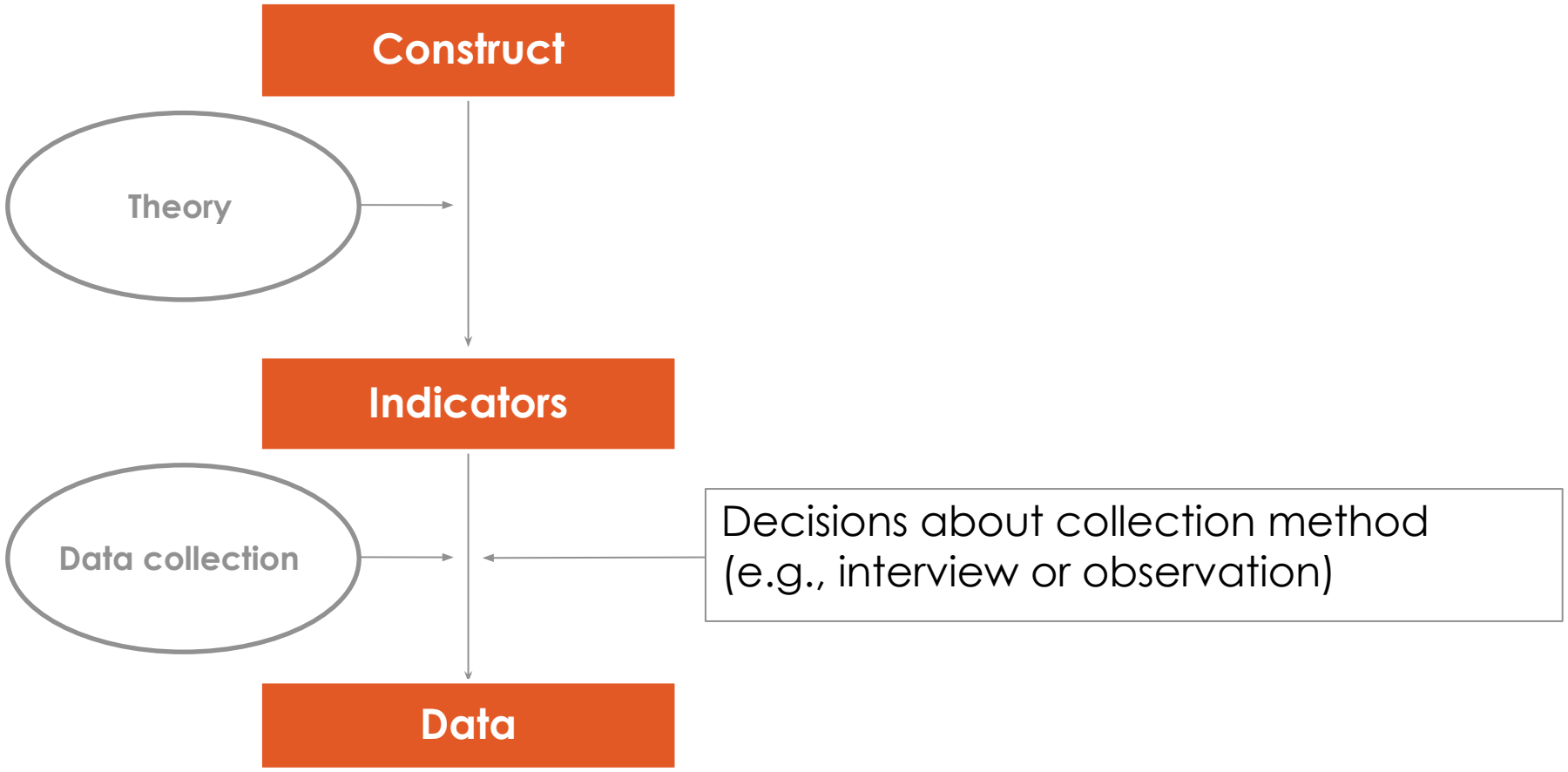
I. Theory of Change

II. Measurement concepts

III. Sources of data

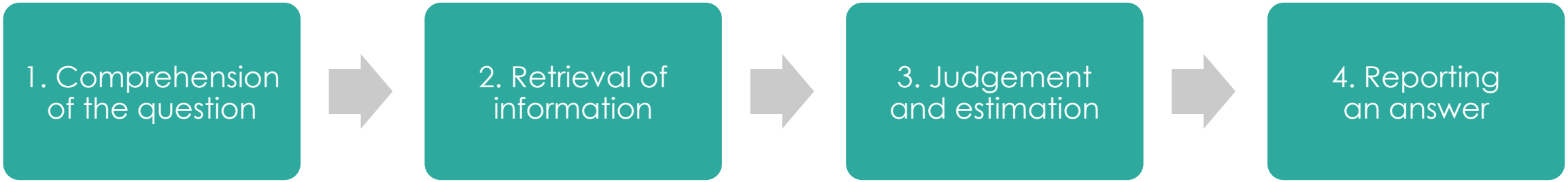
IV. Data collection

Concepts of measurement



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

The response process



Step 1: Comprehension

1. Comprehension of the question → 2. Retrieval of information → 3. Judgement and estimation → 4. Reporting an answer

Step 2: Retrieval

1. Comprehension of the question → **2. Retrieval of information** → 3. Judgement and estimation → 4. Reporting an answer

Step 3: Estimation/ Judgement

1. Comprehension of the question → 2. Retrieval of information → **3. Judgement and estimation** → 4. Reporting an answer

Step 4: Response

1. Comprehension of the question → 2. Retrieval of information → 3. Judgement and estimation → **4. Reporting an answer**

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, and reporting bias

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 outcomes are needed to understand how/why a program did (or did not) work.
- All steps of the Theory of Change should be carefully measured.
- The process of collecting "good" data requires a lot of effort and sometimes involves tradeoffs.
  - Quality vs cost
  - Validity vs reliability

References

Cooper, Ryan, Joseph Doyle, and Andrés Hojman. 2023. "Effects of Enhanced Legal Aid in Child Welfare: Evidence from a Randomized Trial of Mi Abogado". NBER Working Paper No. 30974.

Finkelstein, Amy, Joel C. Cantor, Jesse Gubb, Margaret Koller, Aaron Truchil, Ruohua Annetta Zhou, and Joseph Doyle. 2023. "The Camden Coalition Care Management Program Improved Intermediate Care Coordination: A Randomized Controlled Trial." *Health Affairs* Vol 43, No. 1.

Gubb, Jesse, Cordelia Know, and Matthew Notowidigdo. 2024. "Embracing Failure In Health Care Delivery And Learning From Null Results", Health Affairs Forefront, January 24, 2024

J-PAL. "Health Care Hotspotting in the United States." J-PAL Evaluation Summary.

Resources for further reading

J-PAL Research Resource: Introduction to measurement and indicators

J-PAL Research Resource: Repository of measurement and survey design resources

J-PAL Blog Post: So, you got null results now what?

Reuse and citation

To reference this lecture, please cite as:

J-PAL. 2024. "Lecture: Theory of Change and Measurement." Abdul Latif Jameel Poverty Action Lab. Cambridge, MA.