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
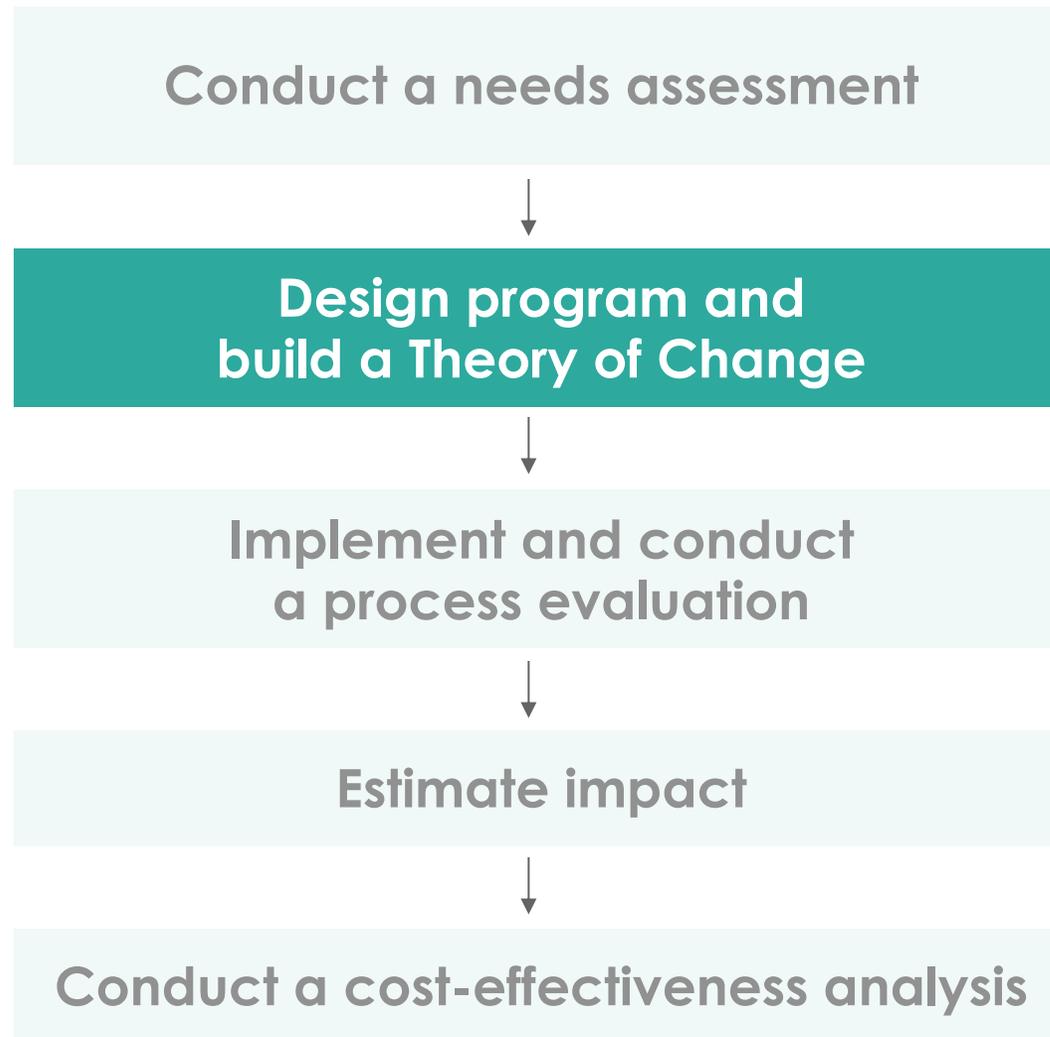10. The Generalizability Framework

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

Conduct a needs assessment

↓

Design program and build a Theory of Change

↓

Implement and conduct a process evaluation

↓

Estimate impact

↓

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?

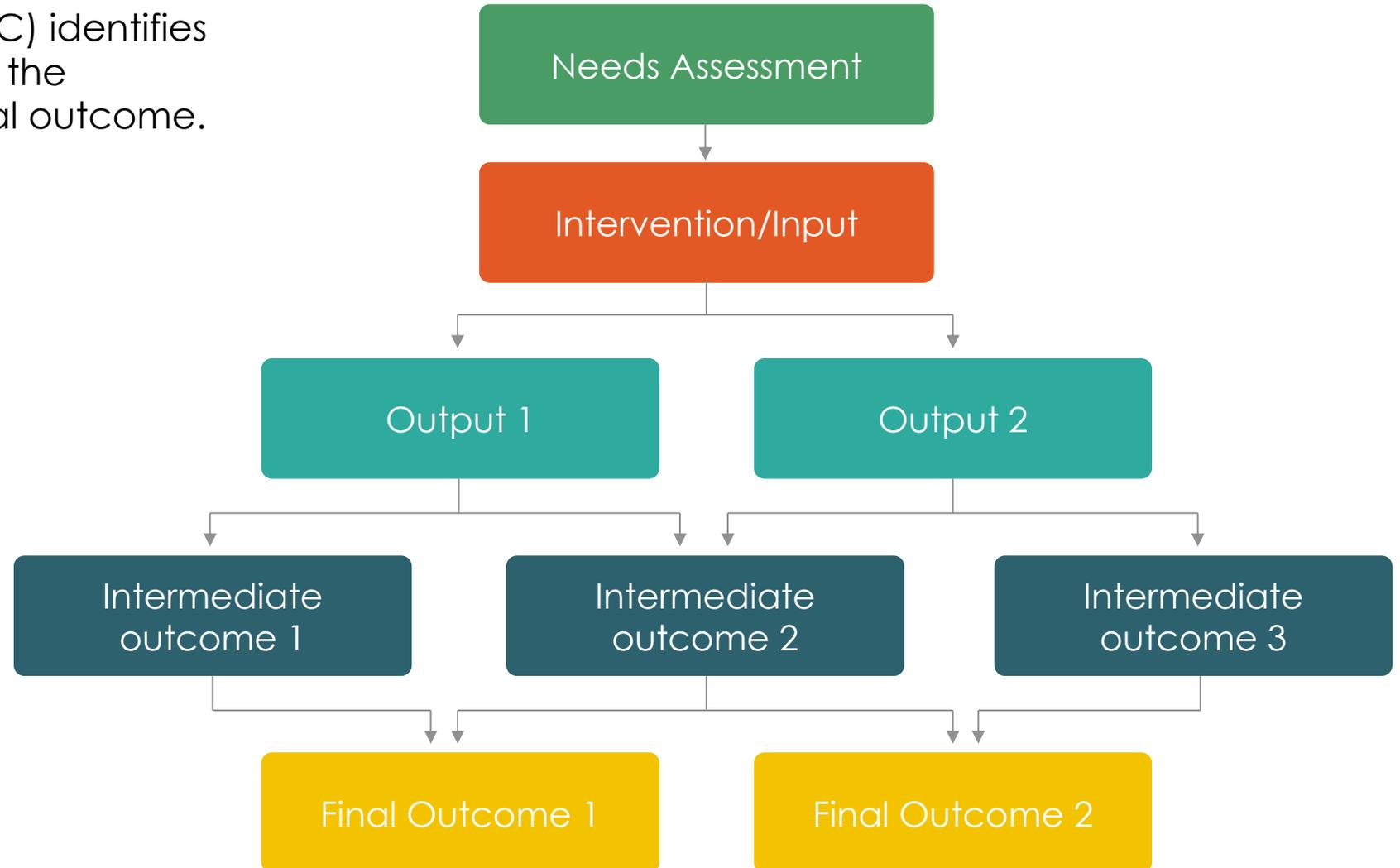**What steps are needed for the program to achieve the desired change in outcomes?**

- What assumptions need to hold?

Input → Output → Intermediate Outcomes → Final Outcomes

Theory of Change

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

Theory of Change
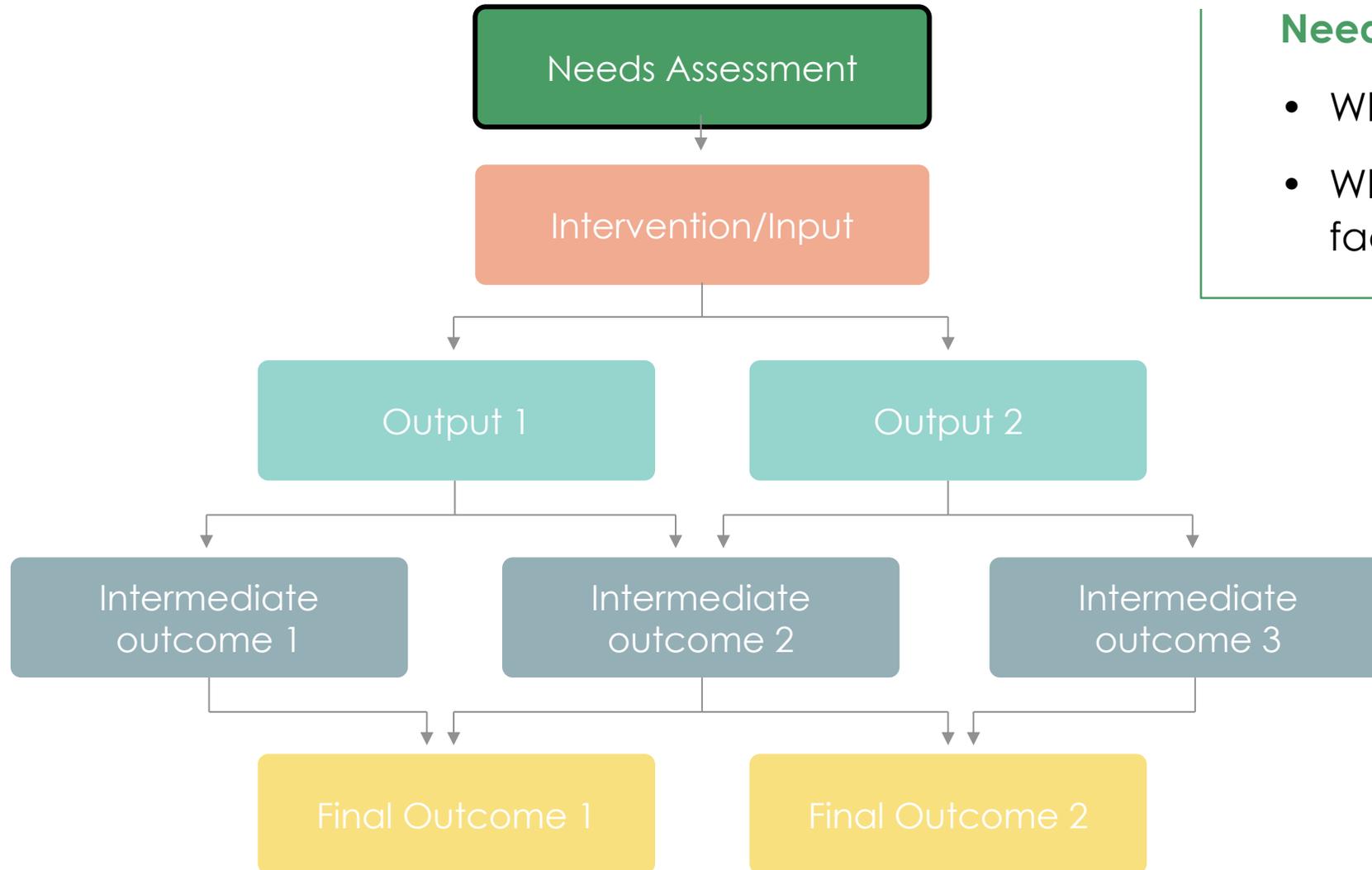

Needs Assessment

Intervention/Input

Output 1

Output 2

Intermediate outcome 1

Intermediate outcome 2

Intermediate outcome 3

Final Outcome 1

Final Outcome 2

Needs Assessment

- What is the core problem/need?
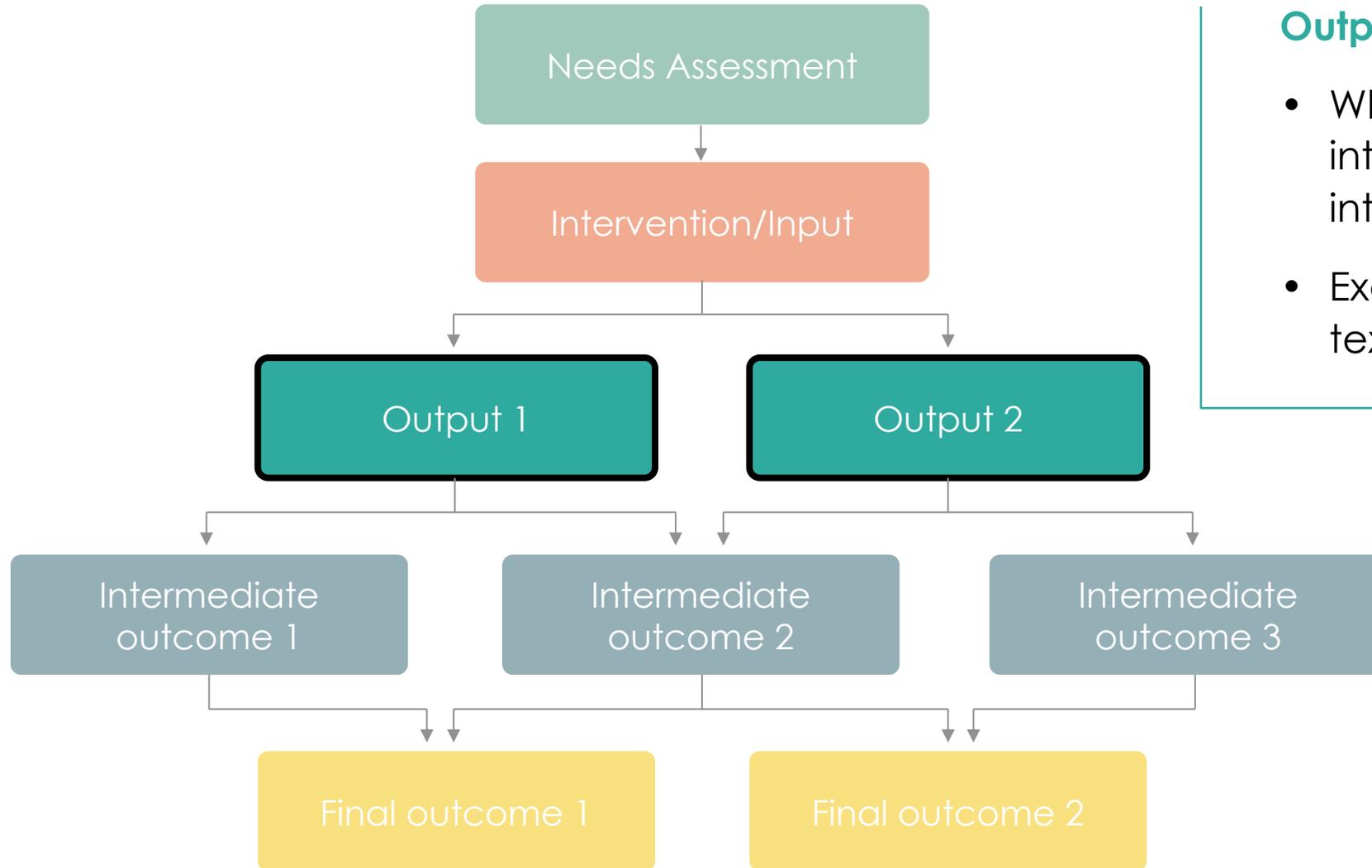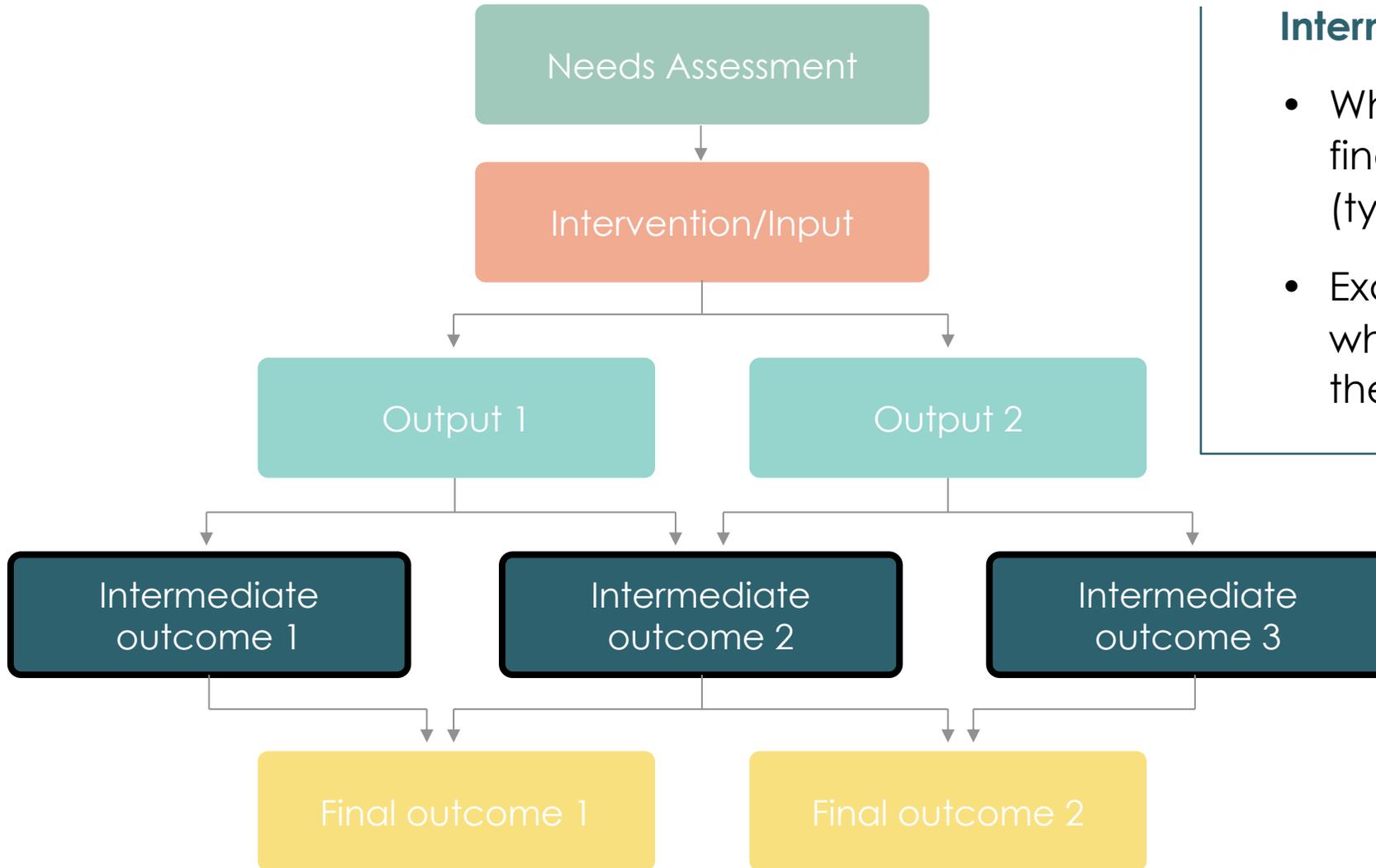
- What are the contributing factors?
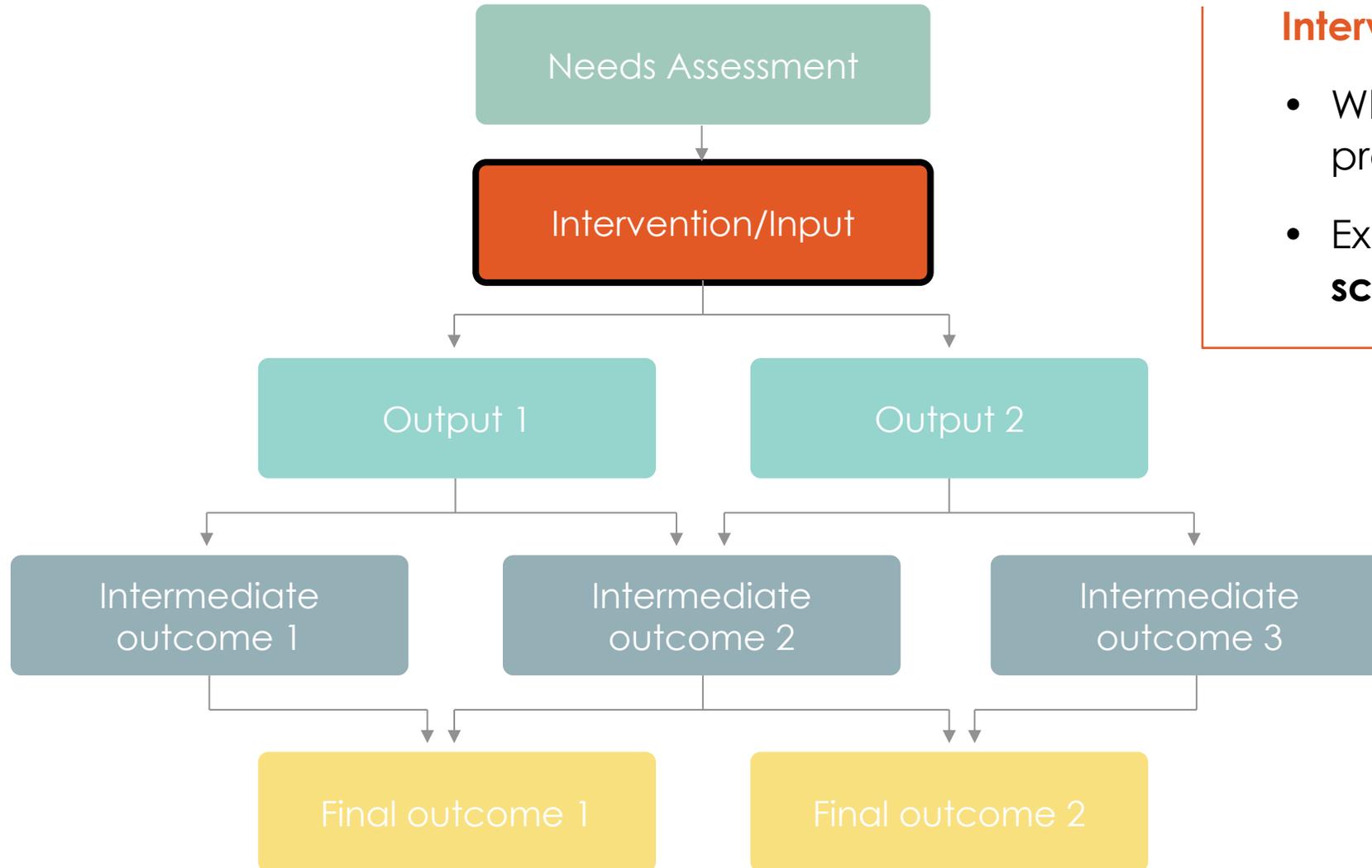
Theory of Change

Output

- What is needed in order for the intervention to be delivered as intended?

- Example: **Students** receive textbooks from schools

Theory of Change


```
Needs Assessment
        |
        v
Intervention/Input
        |
    +---+---+
    |       |
    v       v
Output 1   Output 2
```

Intermediate outcome 1 Intermediate outcome 2 Intermediate outcome 3

Final outcome 1   Final outcome 2

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
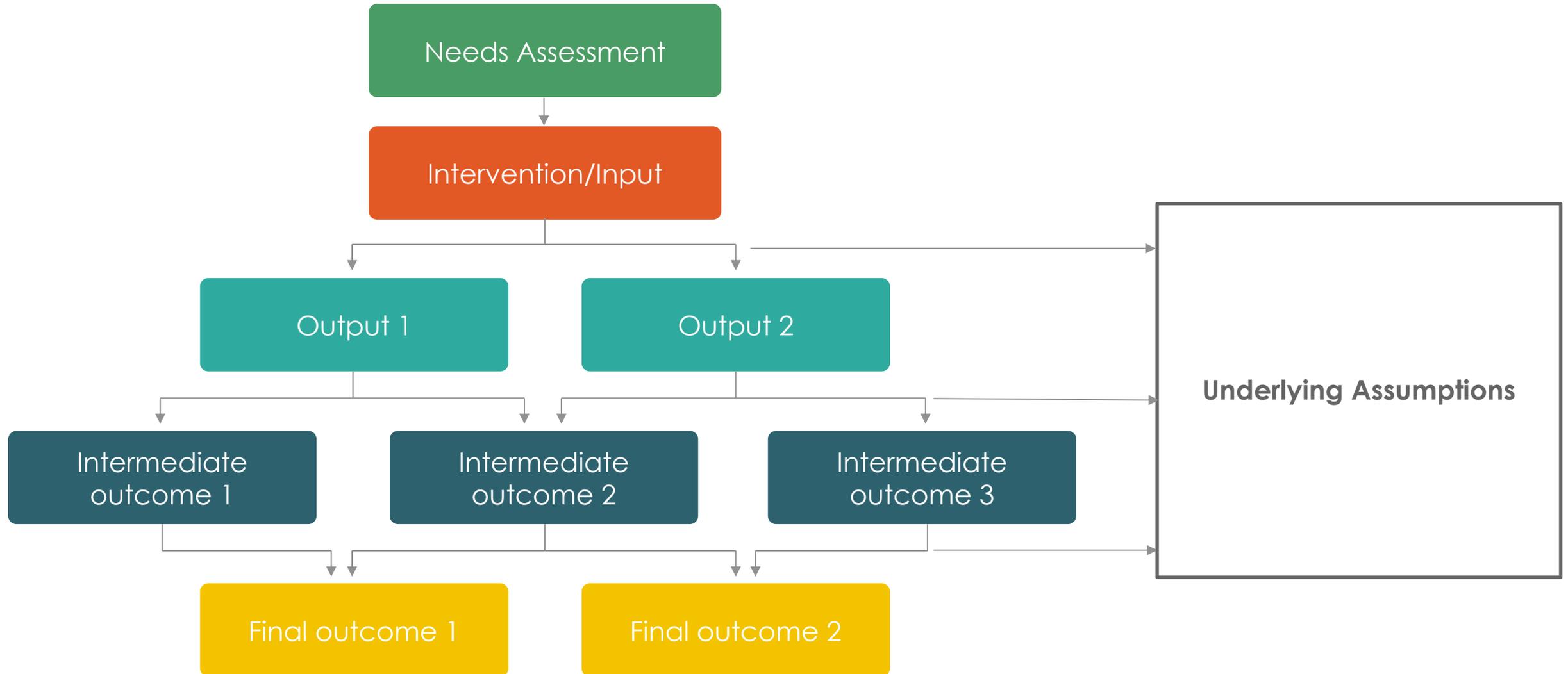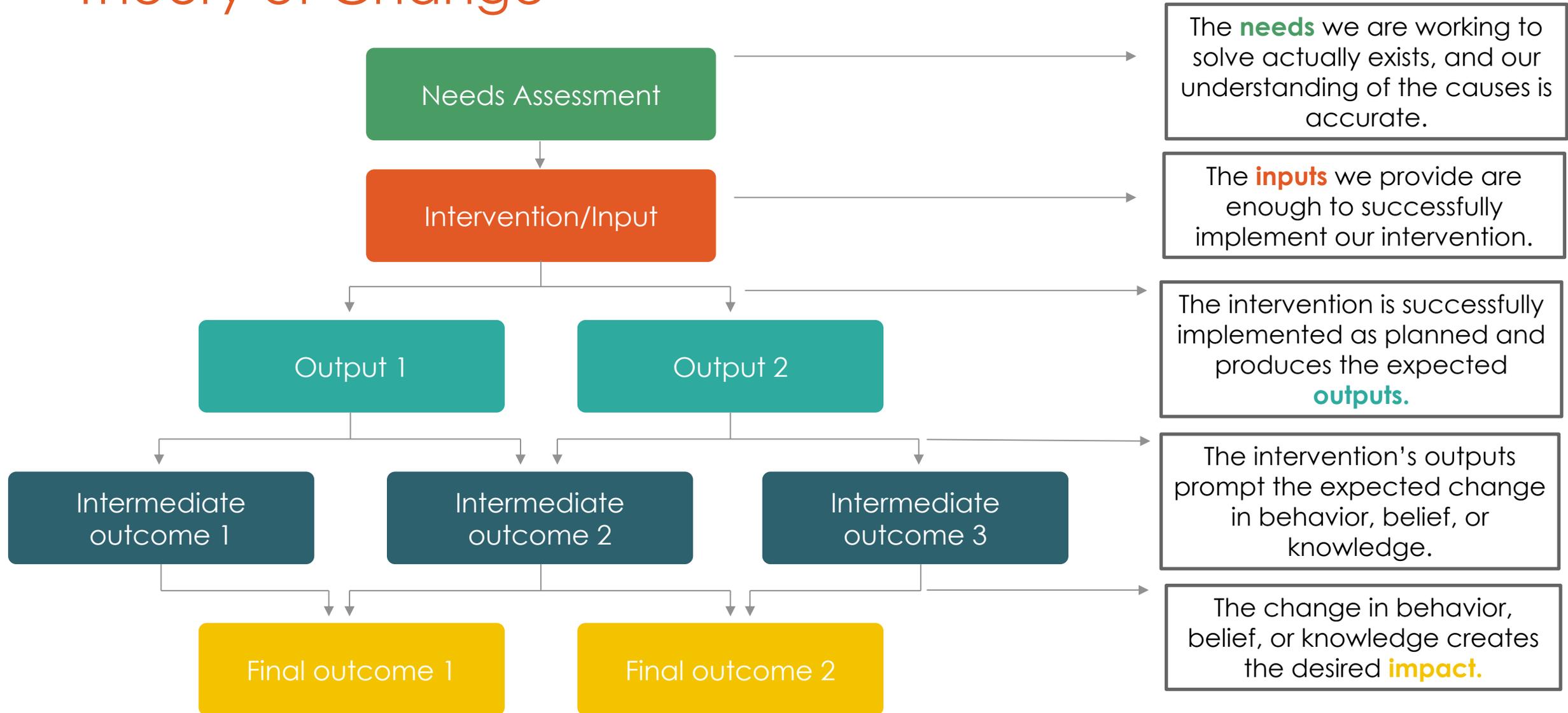
Theory of Change
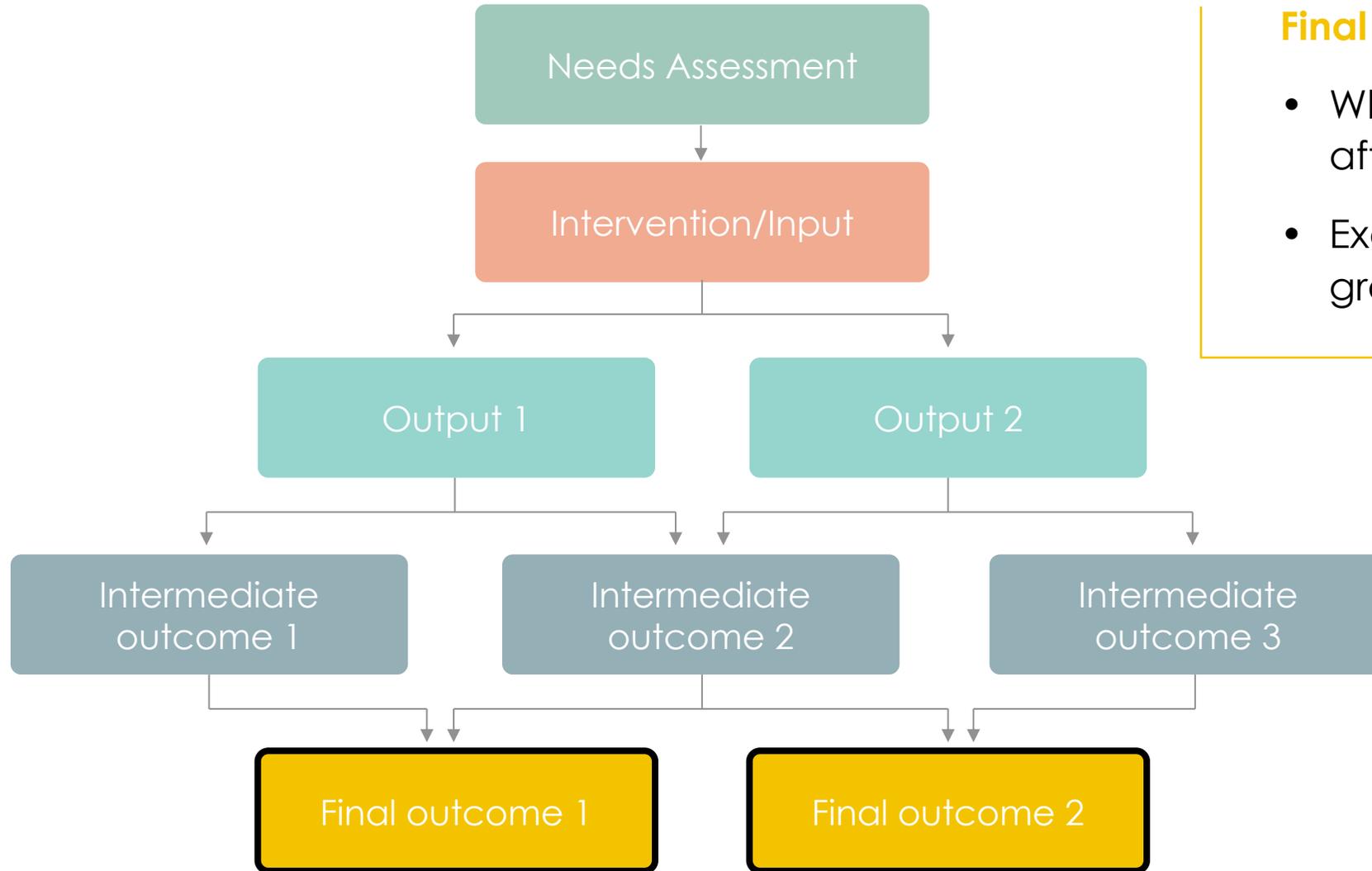
Final outcomes

- What do you ultimately want to affect with the intervention?

- Example: Increase in high school graduation rates

Needs Assessment

Intervention/Input

Output 1

Output 2

Intermediate outcome 1

Intermediate outcome 2

Intermediate outcome 3

Final outcome 1

Final outcome 2

Theory of Change

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

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.**

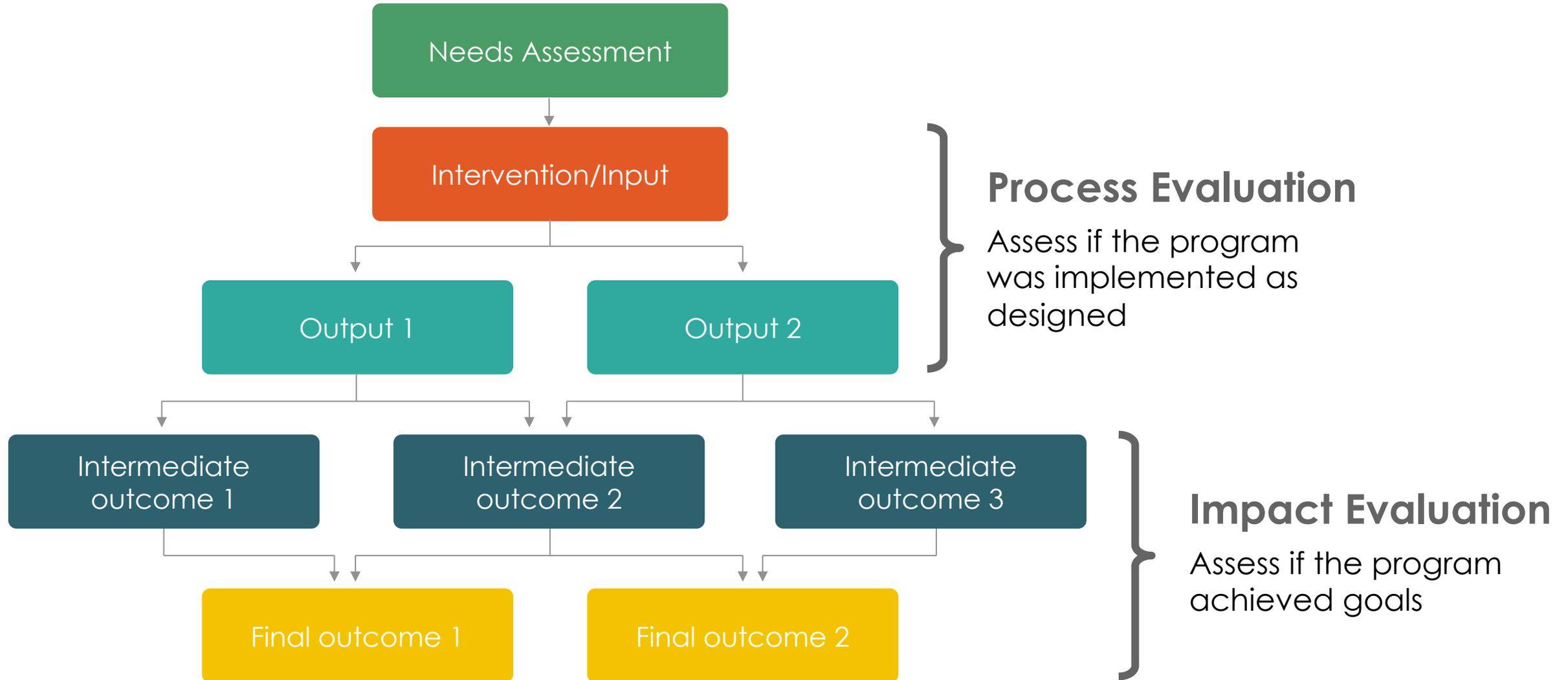# Why spend time on a Theory of Change?

1. **Helps design the 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 | Output 2

Intermediate outcome 1 | Intermediate outcome 2 | Intermediate outcome 3

Final outcome 1 | Final outcome 2

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.
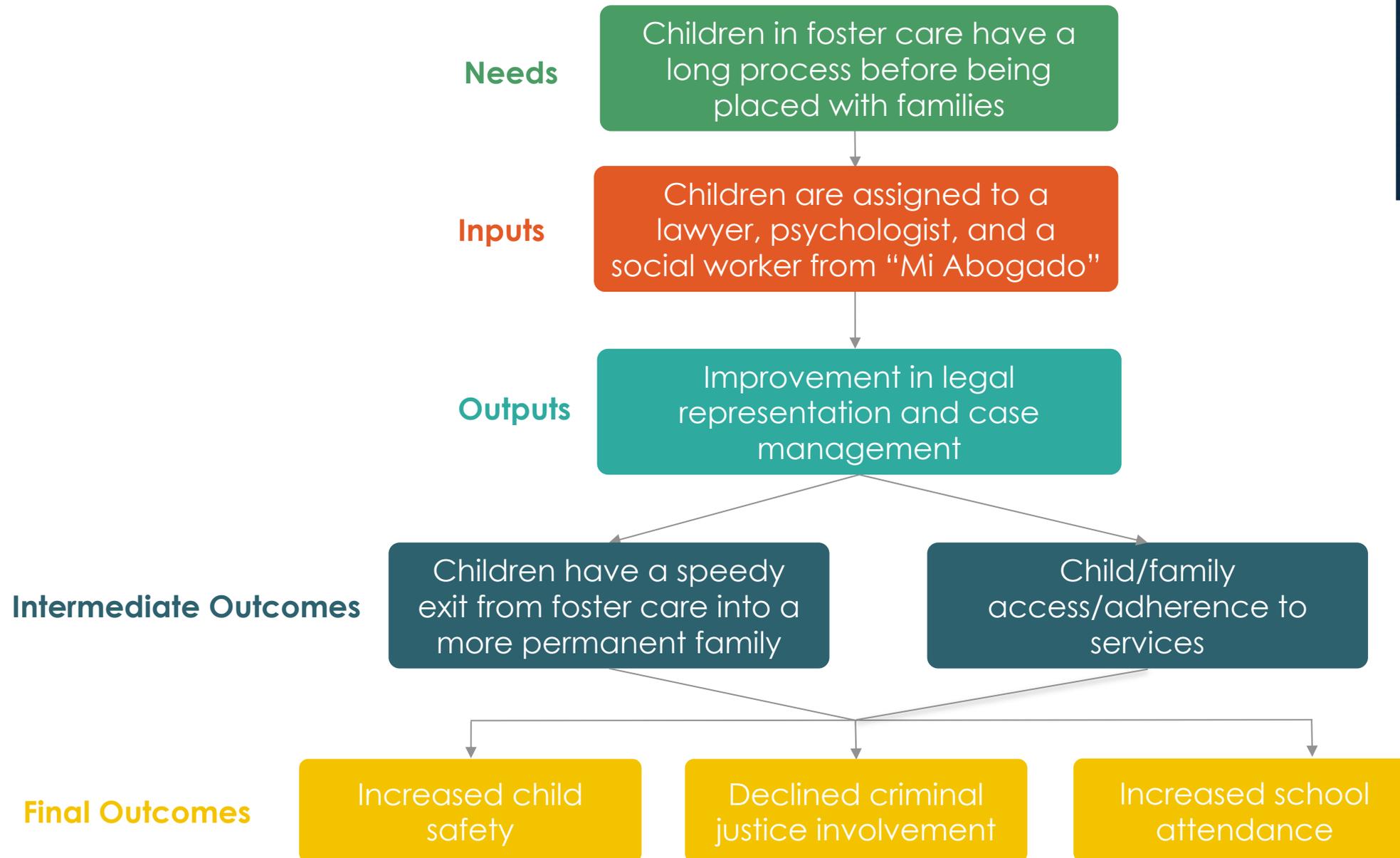
Why spend time on a Theory of Change?

- 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

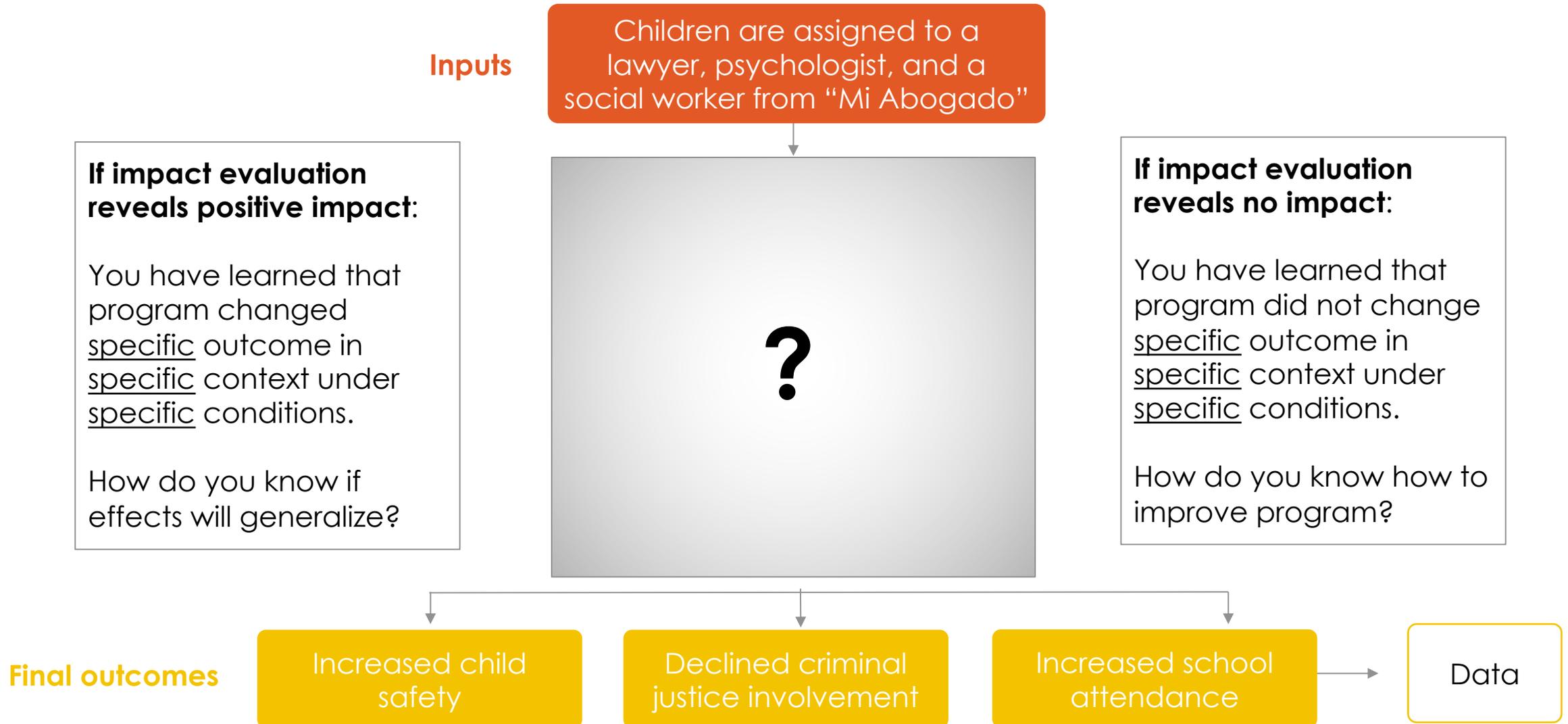 – Deciding which data to collect

Discussion question:

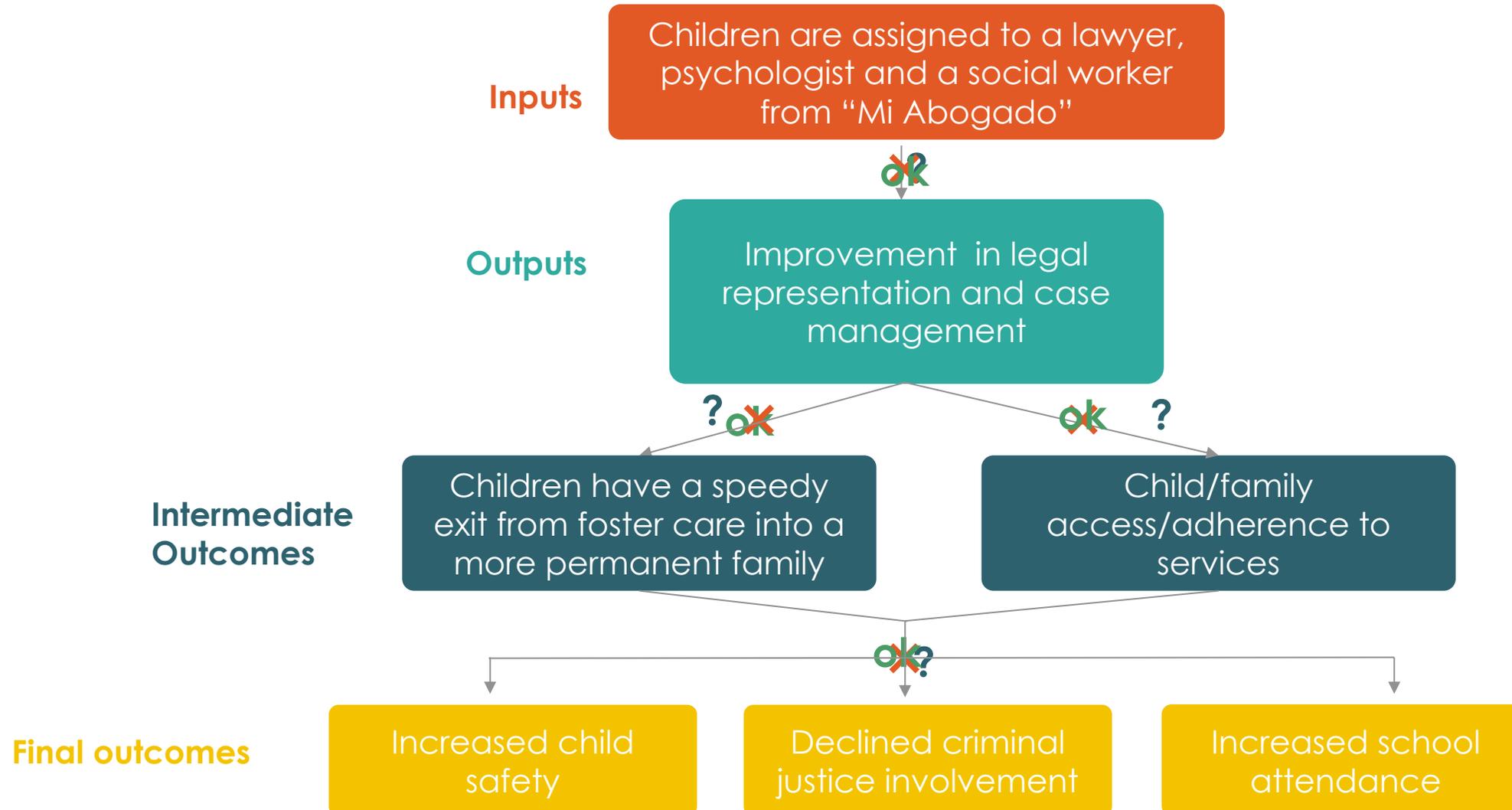What happens if we find no impact?

Discussion question:

What happens if we find no impact?

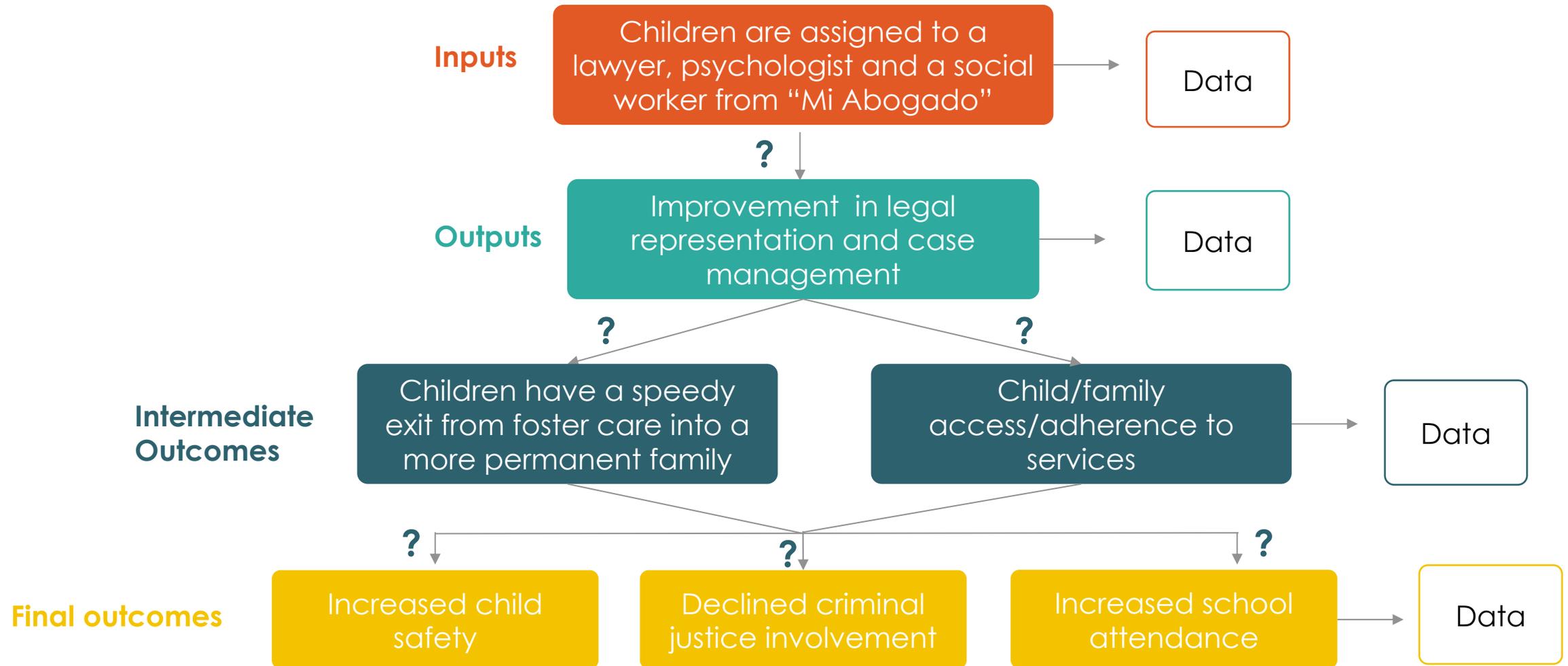What happens if we do find an impact?

Without a Theory of Change: Unknowns

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

If impact evaluation reveals positive impact:

You have learned that program changed <u>specific</u> outcome in <u>specific</u> context under <u>specific</u> conditions.

How do you know if effects will generalize?

?

If impact evaluation reveals no impact:

You have learned that program did not change <u>specific</u> outcome in <u>specific</u> context under <u>specific</u> conditions.

How do you know how to improve program?

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

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

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

Outline

I. Theory of Change

II. Measurement concepts

III. Sources of data

IV. Data collection

Concepts of measurement
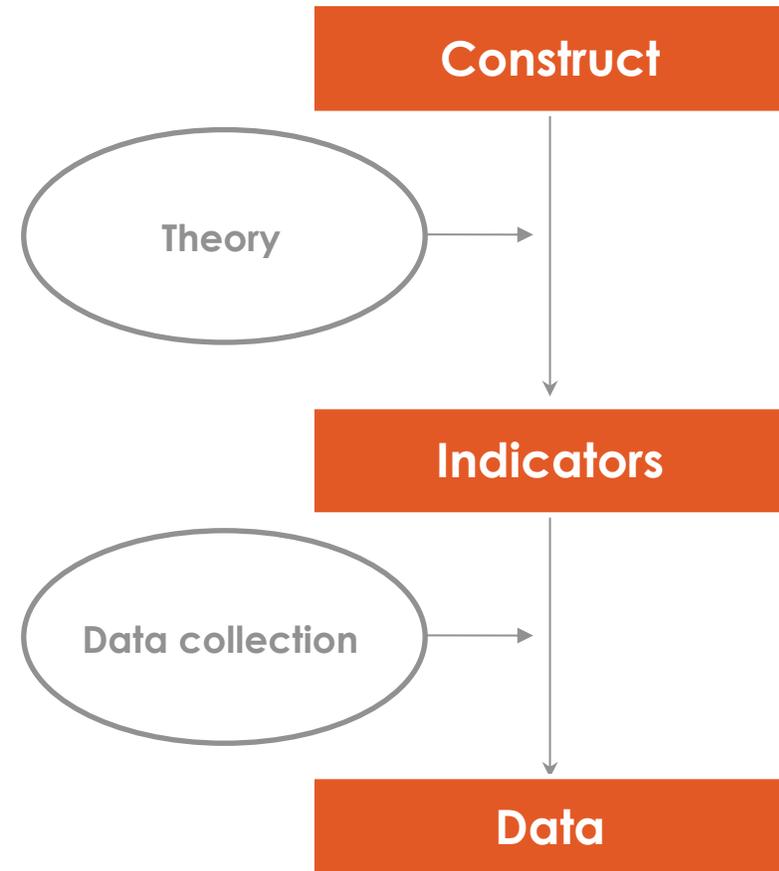
Construct

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

Theory

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

Low   High

- **Reliability** – measuring the thing precisely

Low   High

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

Construct

Theory

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

Indicators

Data collection

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

Data

LOW   HIGH
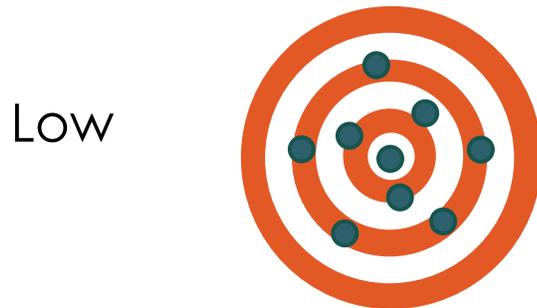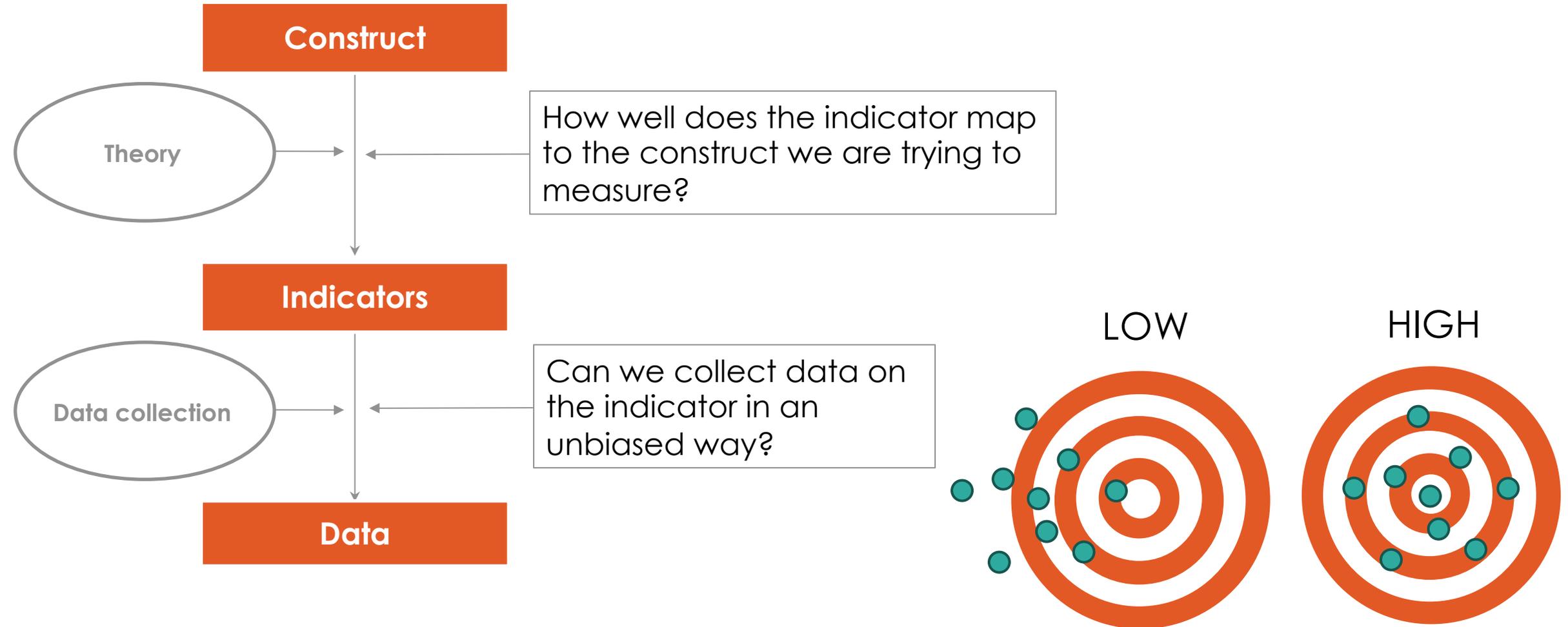
Question:

Construct: Students' school performance

Indicator: School attendance rates

Where does that indicator land?

Validity
(a.k.a. accuracy/unbiased)

LOW   HIGH

A   B

Reliability (a.k.a. precision)

Construct

Theory

How time/situation specific is the indicator?

Indicators

Data collection

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

Data

LOW

HIGH
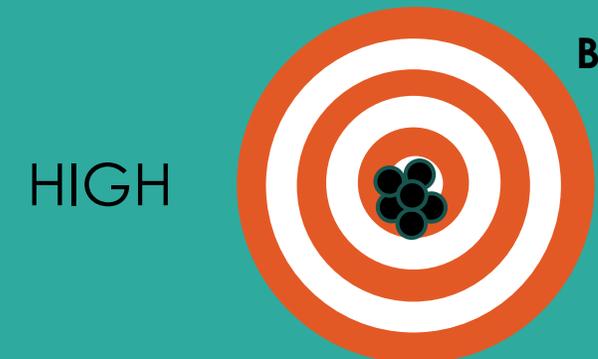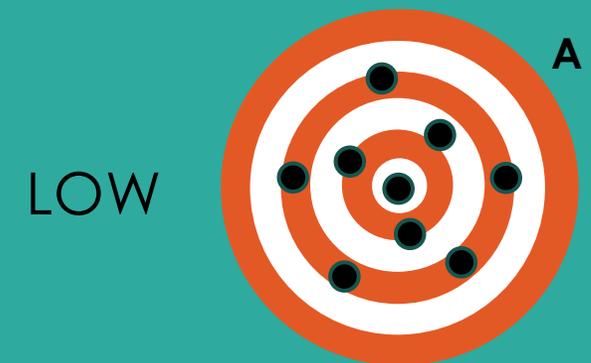
The goals of measurement
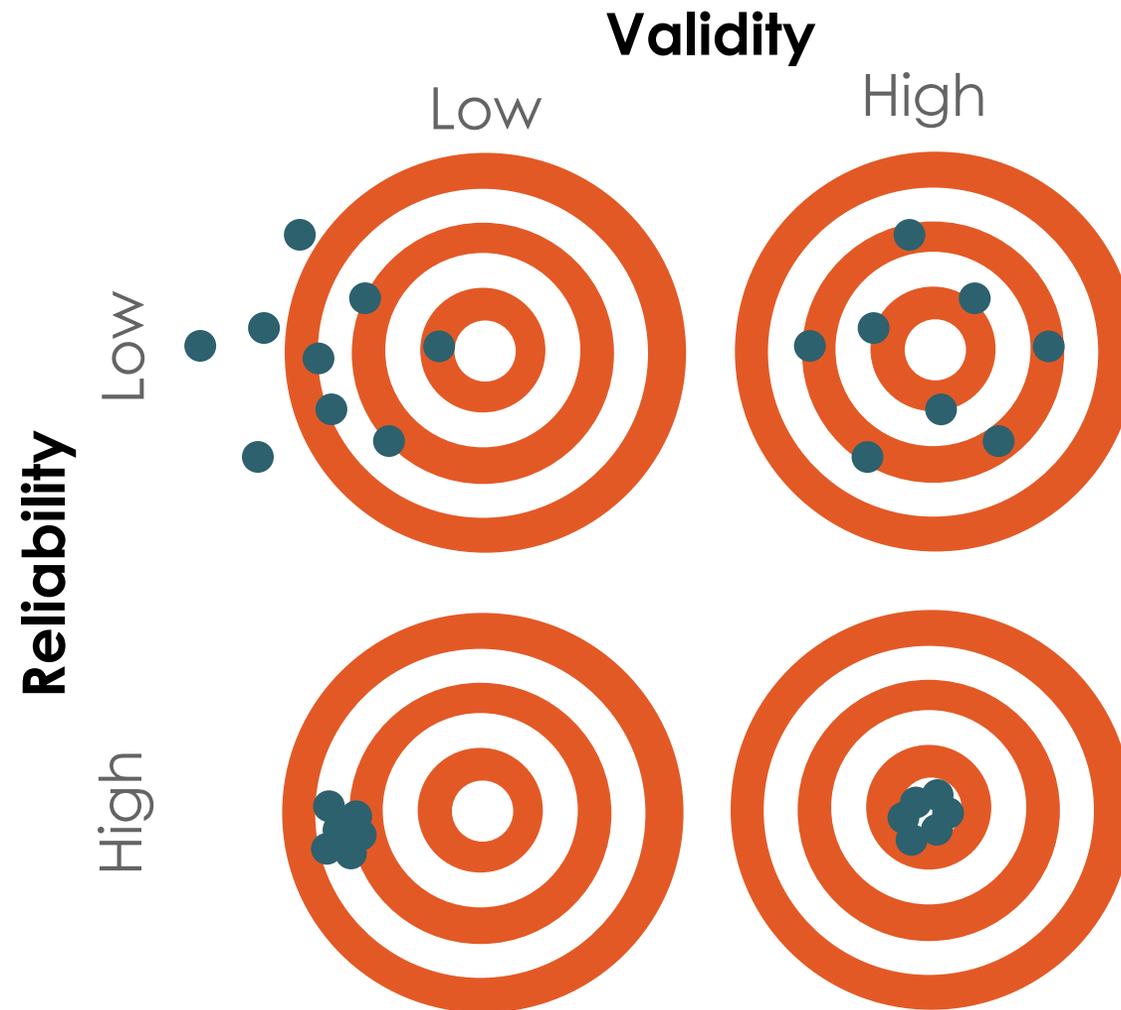
Maximizing validity

Theory: Think about how the indicator maps to the construct

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

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

LOW   HIGH
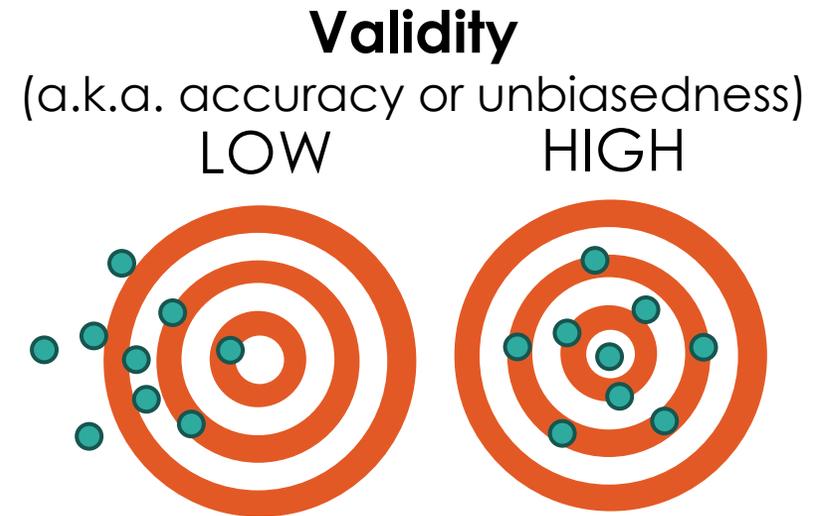
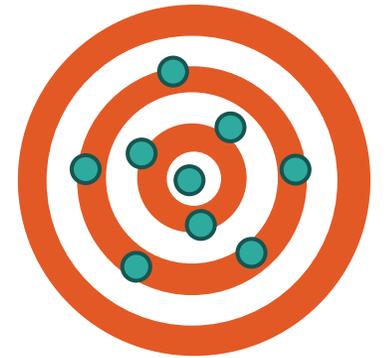| Possible sources of bias:	Possible ways to minimize bias
• Theory/mapping • Experimenter demand effects • Social desirability bias • Recall bias • Translation/interpretation	• Use administrative data where possible • Use methods for collecting data on sensitive topics • Back translation and piloting • Use multiple data sources

Maximizing reliability

LOW

Reliability
(a.k.a.
precision)

HIGH

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:	**Possible ways to maximize reliability**
- Fatigue - Ambiguous wording (e.g. "# people in household")	- Consider survey length - Consider the answer choices - 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
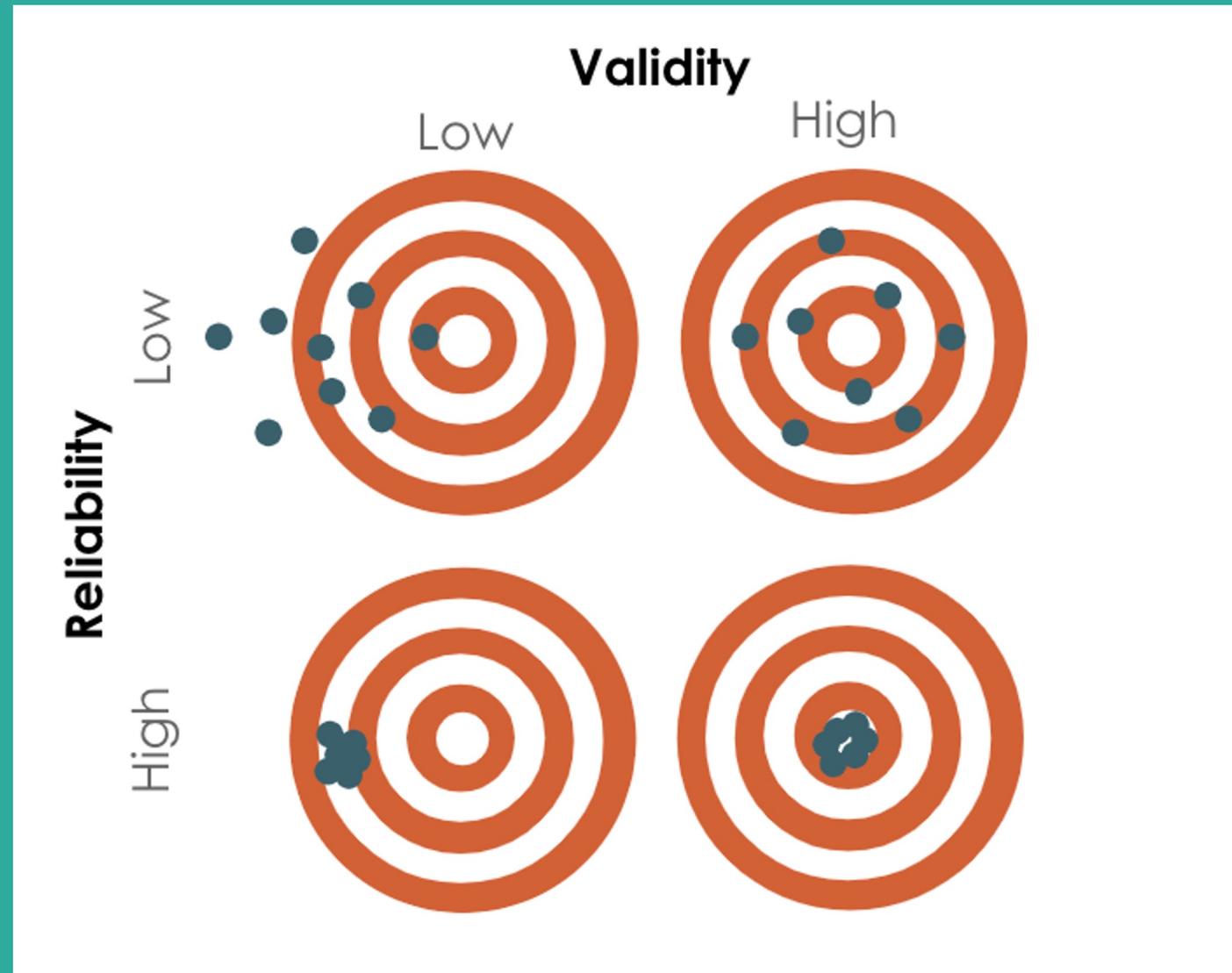
Outline

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	**<u>NOT</u> 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
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?
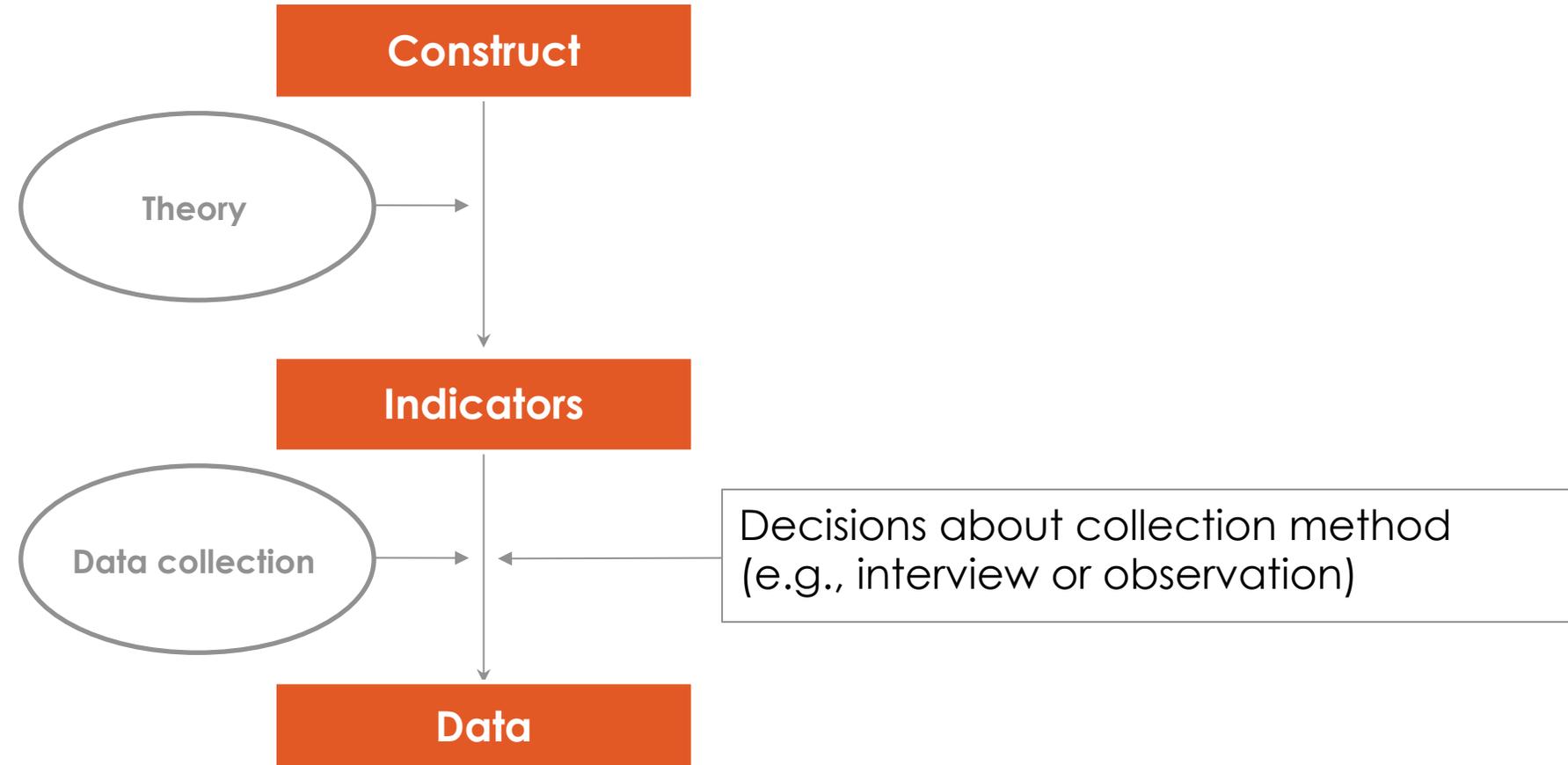
A. Administrative data

B. Survey data

C. Both

D. Other

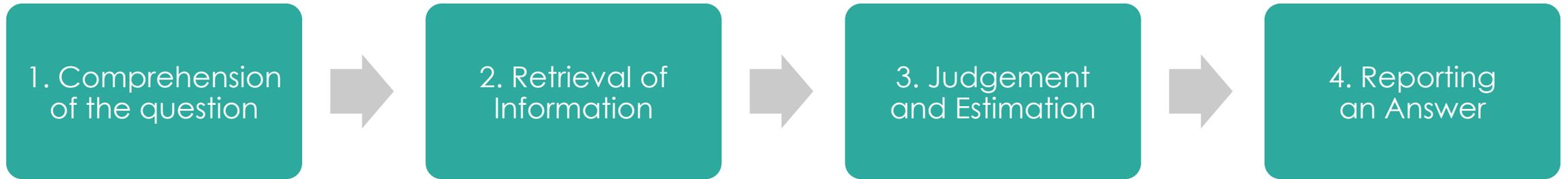# Outline

I.   Theory of Change

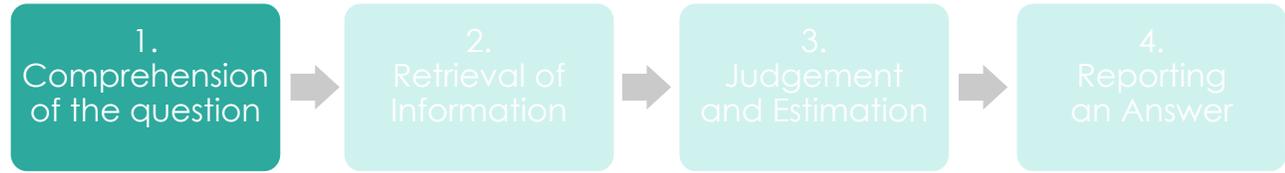II. Measurement concepts

III. Sources of data

IV. Data collection

Concepts of measurement
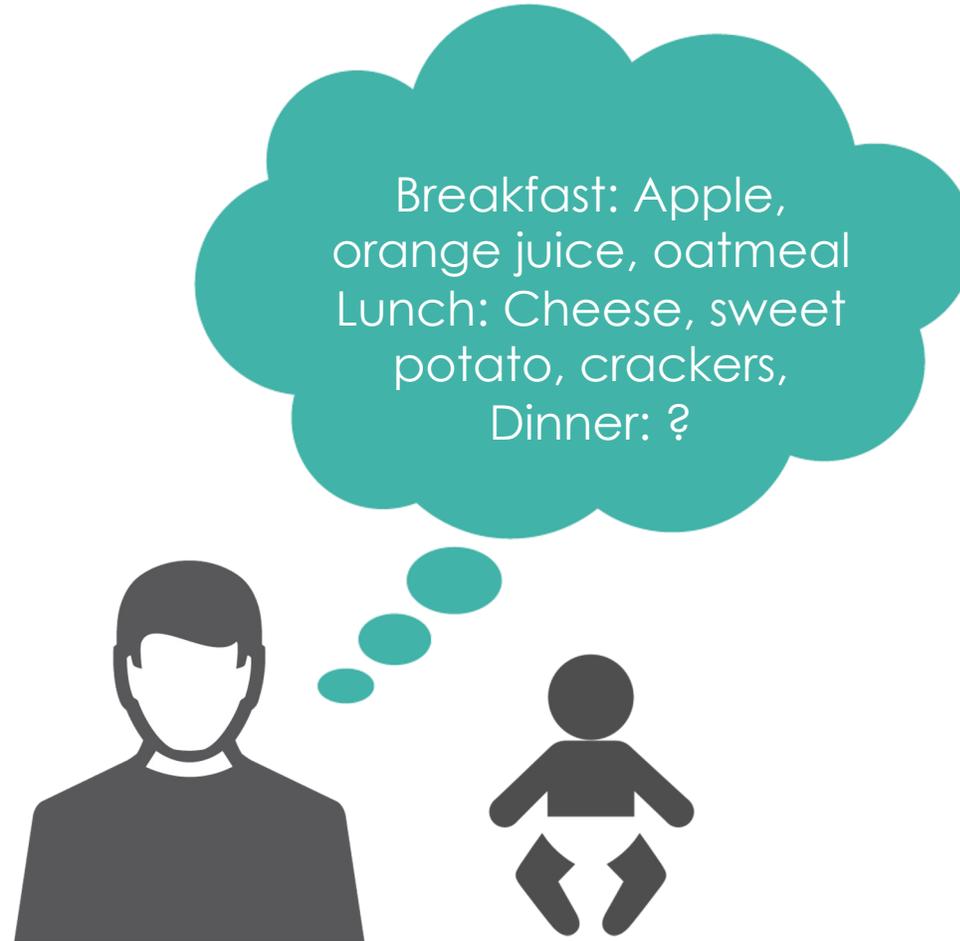

Construct
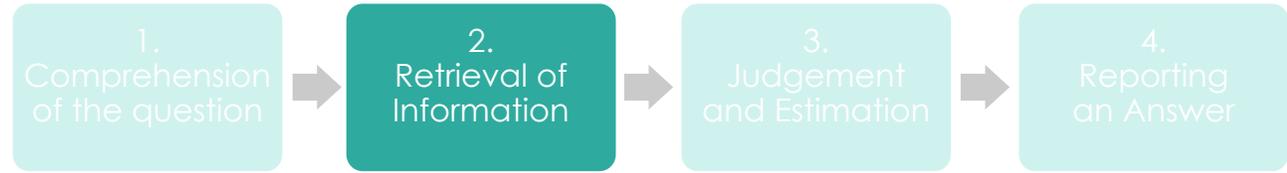
Theory

Indicators

Data collection

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

Data

The Response Process

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

Step 2: Retrieval

Step 3: Estimation/ Judgement

2 fruits, 0 vegetables at breakfast, 0 fruits, 1 vegetable at lunch… Maybe 2 vegetables at dinner?
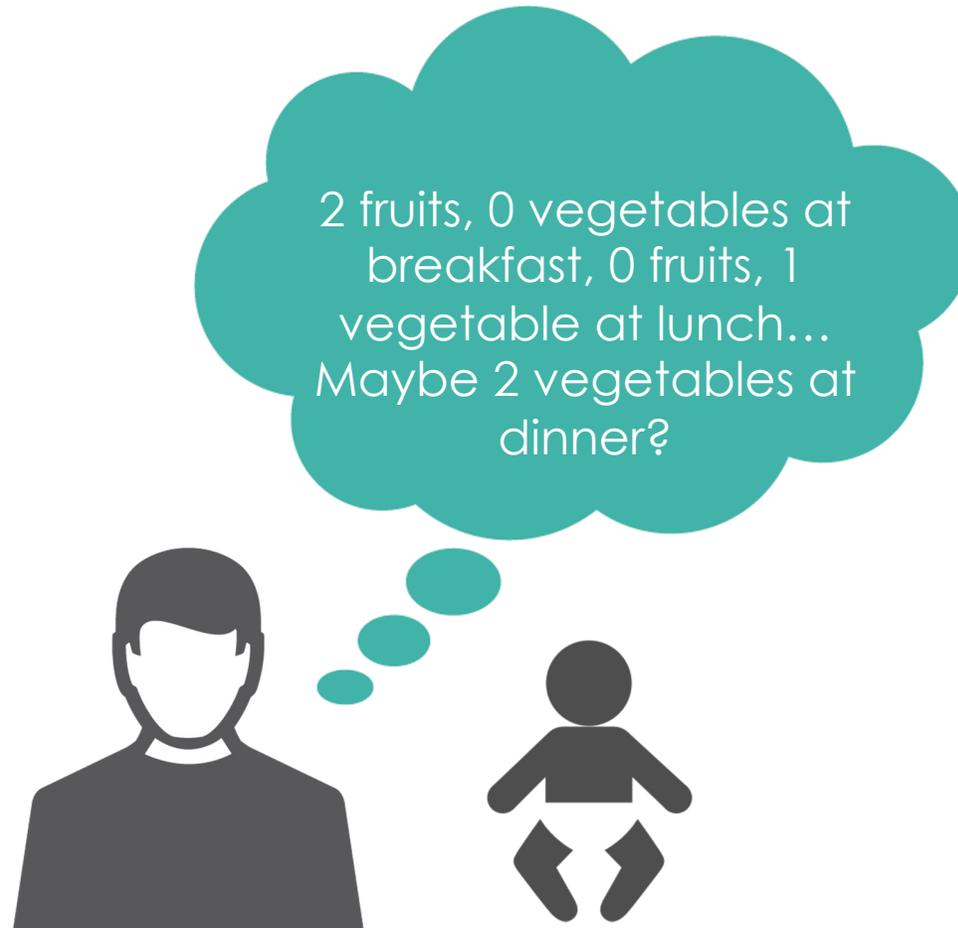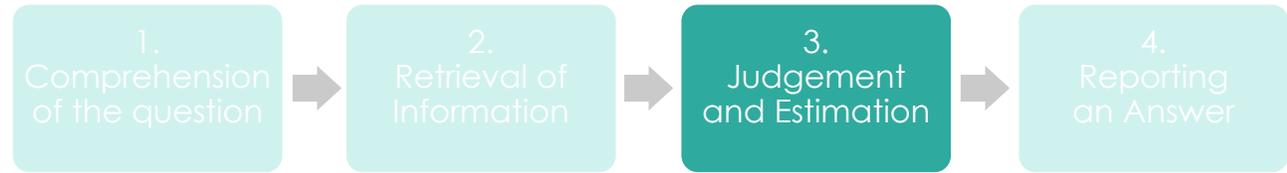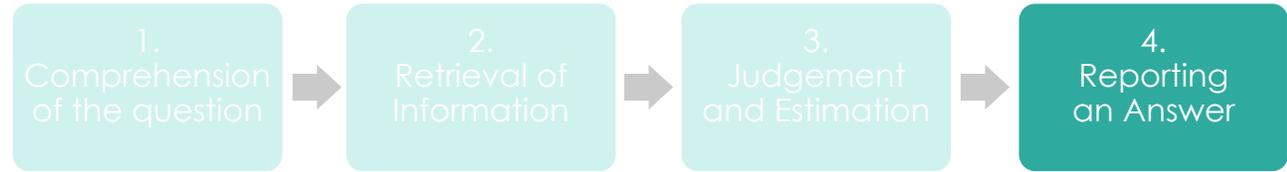
Summary: Potential sources of measurement error

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

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

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

Cooper, Ryan and Doyle, Joseph John and Hojman, Andrés, Effects of Enhanced Legal Aid in Child Welfare: Evidence from a Randomized Trial of Mi Abogado (February 2023). NBER Working Paper No. w30974, Available at SSRN: https://ssrn.com/abstract=4364722

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:

J-PAL. "Lecture: Theory of Change and Measurement." Abdul Latif Jameel Poverty Action Lab. 2023. Cambridge, MA.

 J-PAL, 2023