

# Welcome!

- The program will begin shortly
- Please turn on your video
- Please mute your microphone
- You are welcome to use the chat box for discussions or questions throughout the session!
- Please rename your screen to
  - “Name, Preferred Pronouns, State/City/County”
  - *Example: **Jennifer Lopez, She/Her, The Bronx, NY***



# Building on Theory of Change: Measurement and Process Evaluation

J-PAL North America  
Kim Dadisman and Andrea Salas  
August 16, 2023

# Workshop learning objectives

- Build upon learnings from the April convening.
- Strengthen knowledge about measurement and its relation to program evaluation and the theory of change.
- Build capacity to engage stakeholders on evaluation and measurement.
- Learn more about process evaluation and its importance for impact evaluation.
- Through this session, Fellows will understand their next steps and develop key action items.

# Trainers



**Kim Dadisman**  
Associate Director of  
Policy



**Andrea Salas**  
Senior Policy & Training  
Associate

# Agenda

- I. ToC and stakeholder engagement
- II. Measurement recap and deeper dive
- III. Process evaluation
- IV. Q&A



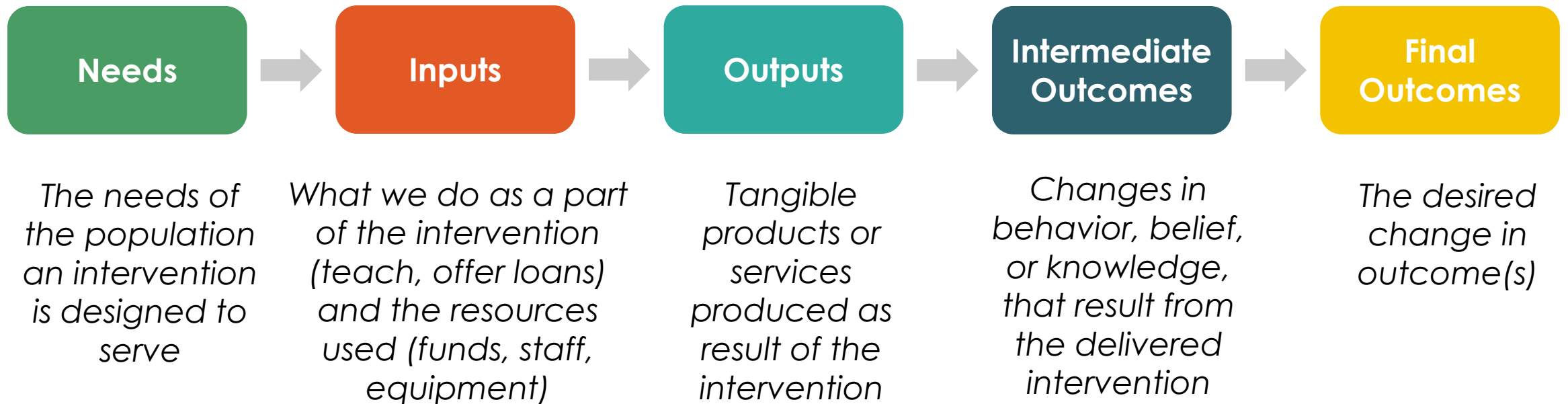
# Agenda

- I. ToC and stakeholder engagement
- II. Measurement recap and deeper dive
- III. Process evaluation
- IV. Q&A

# What is a theory of change?

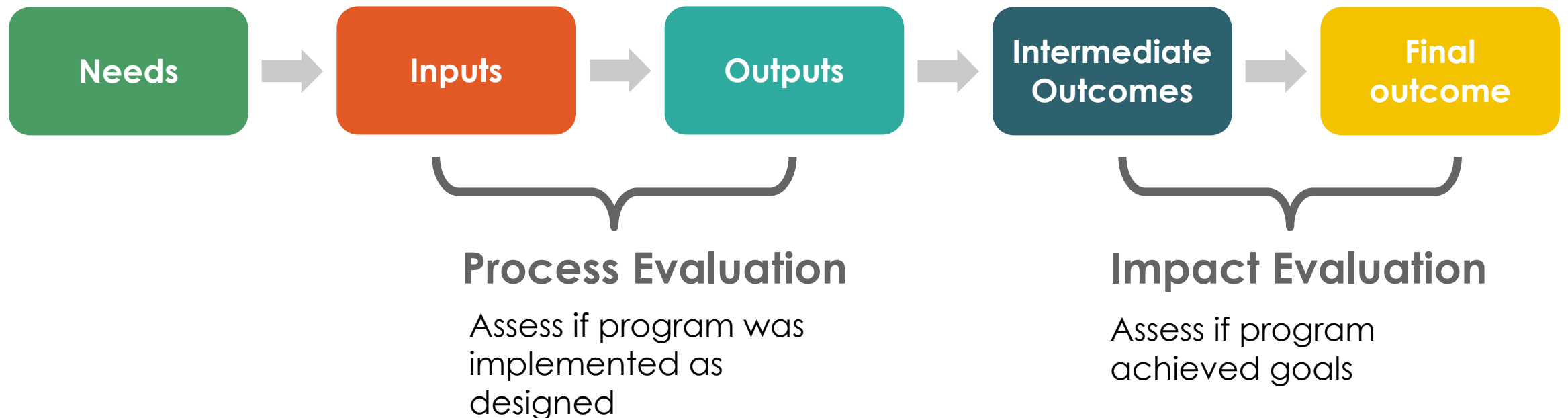
A theory of change (ToC) is a comprehensive description of how and why desired change is expected to happen in a particular context.

**Simply put, it is a road-map for your program**



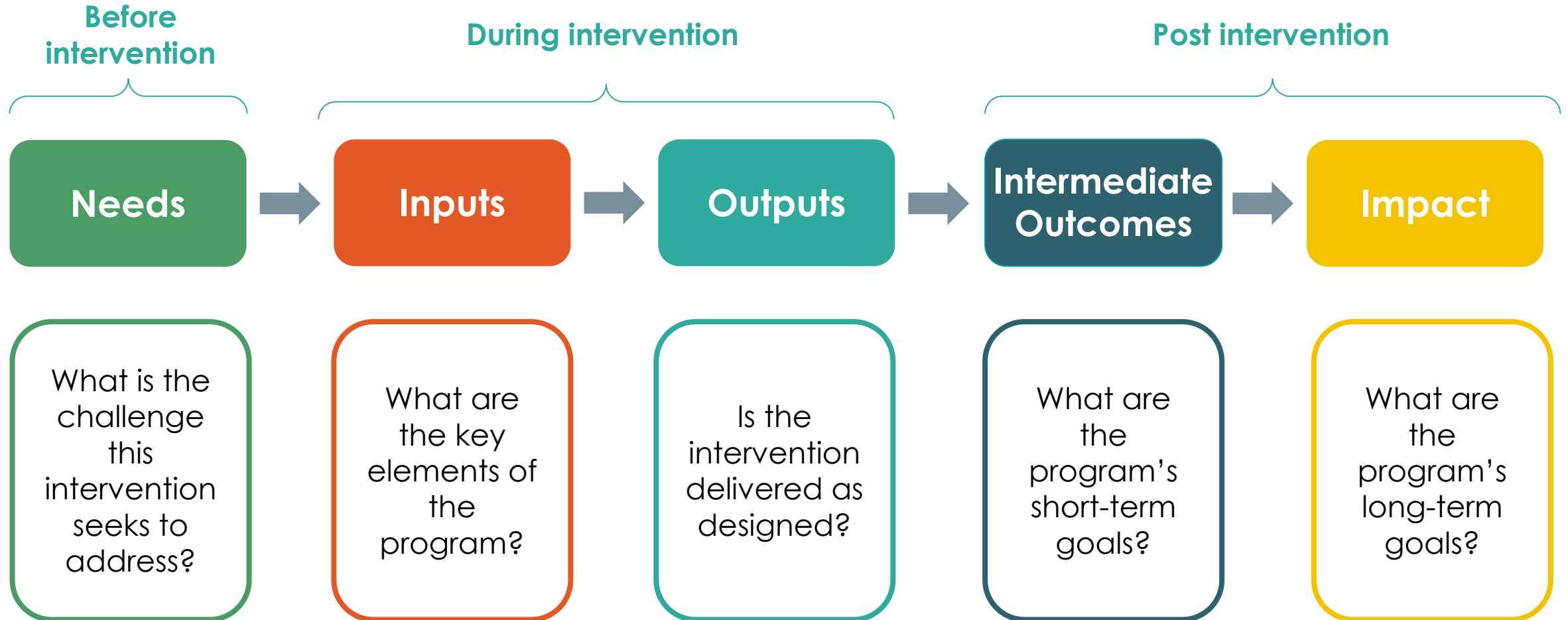
# Measurement along a theory of change

- We use a theory of change to guide **process evaluations** to monitor if the program is implemented as planned
- We also use a theory of change to identify the research question(s) for an **impact evaluation**





# Communication with stakeholders across the theory of change is key



# Identifying key stakeholders

- It “takes a village” to implement a program and evaluate it
- To identify who is (and should be) in your “village,” it could be helpful to...



Conduct a landscape analysis with asset mapping

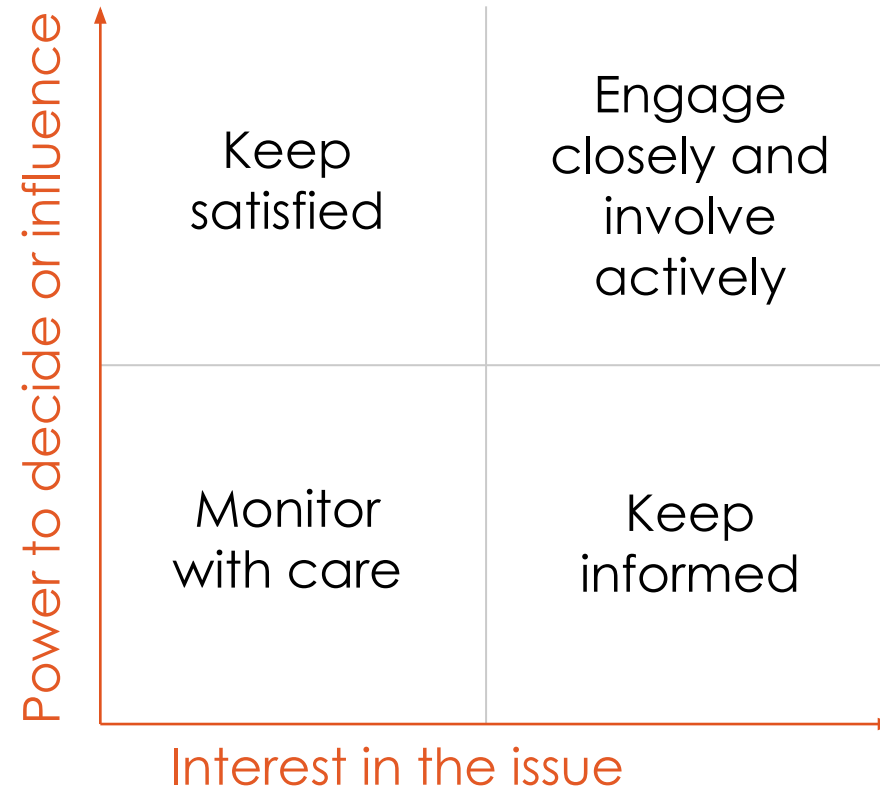


Map your program ecosystem



Look around the “table”

# How do you balance the interests and needs from multiple stakeholders?



Based on Chart 10-5, Stakeholder Analysis Matrix ©2008 Project Management Institute. A Guide to the Project Management Body of Knowledge®, (PMBOK® Guide) – Fourth Edition

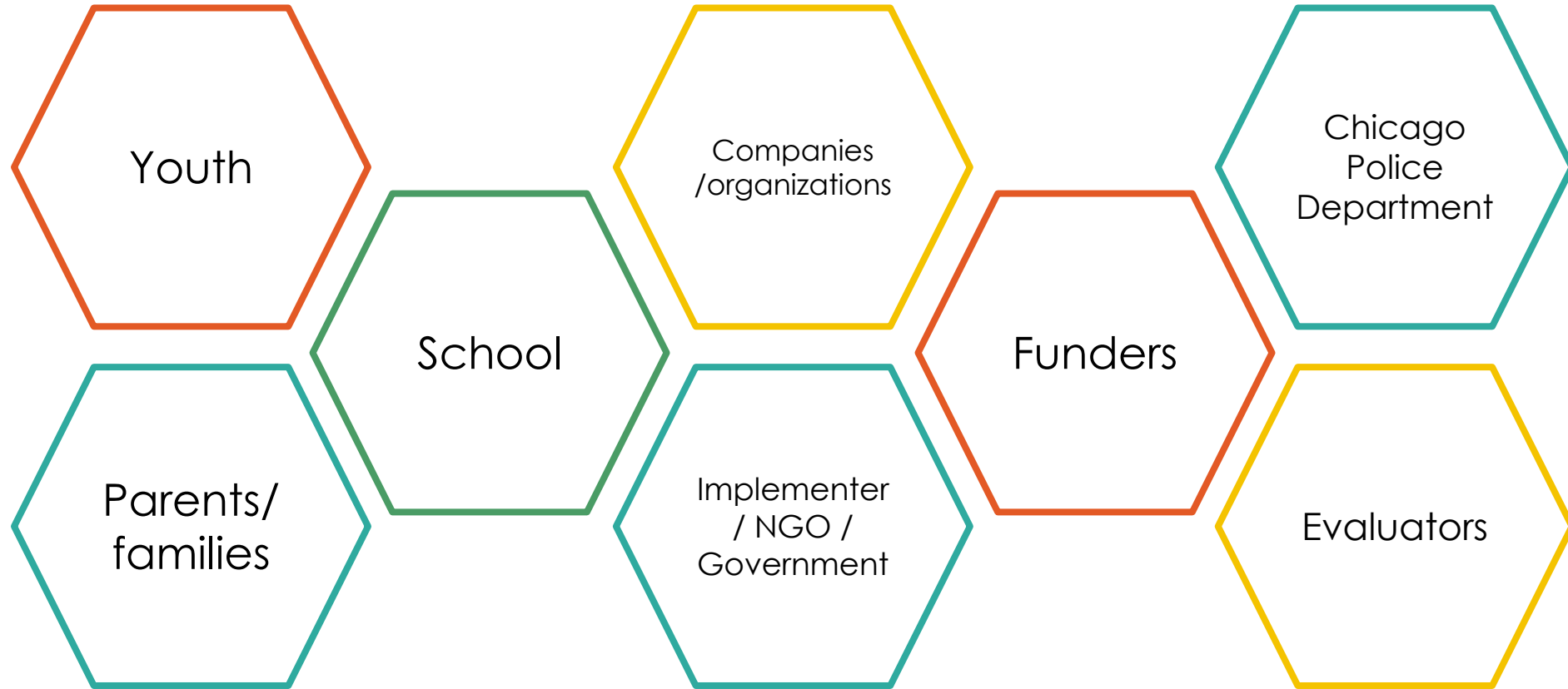
# Summer Youth Employment Programs (SYEP)



Photo: University of Chicago Crime Lab

Based on "Summer Jobs Reduce Violence among Disadvantaged Youth" by Sara B. Heller (2014), *Science*; "Stopping a Bullet With a Summer Job," J-PAL; and *J-PAL Voices: The Impact and Promise of Summer Jobs in the United States*, J-PAL.

# SYEP potential stakeholders



Who are the key stakeholders that you engage with?

Before we move on:

**Are there any questions?**



# Agenda

- I. ToC and stakeholders engagement
- II. Measurement recap and deeper dive
- III. Process evaluation
- IV. Q&A



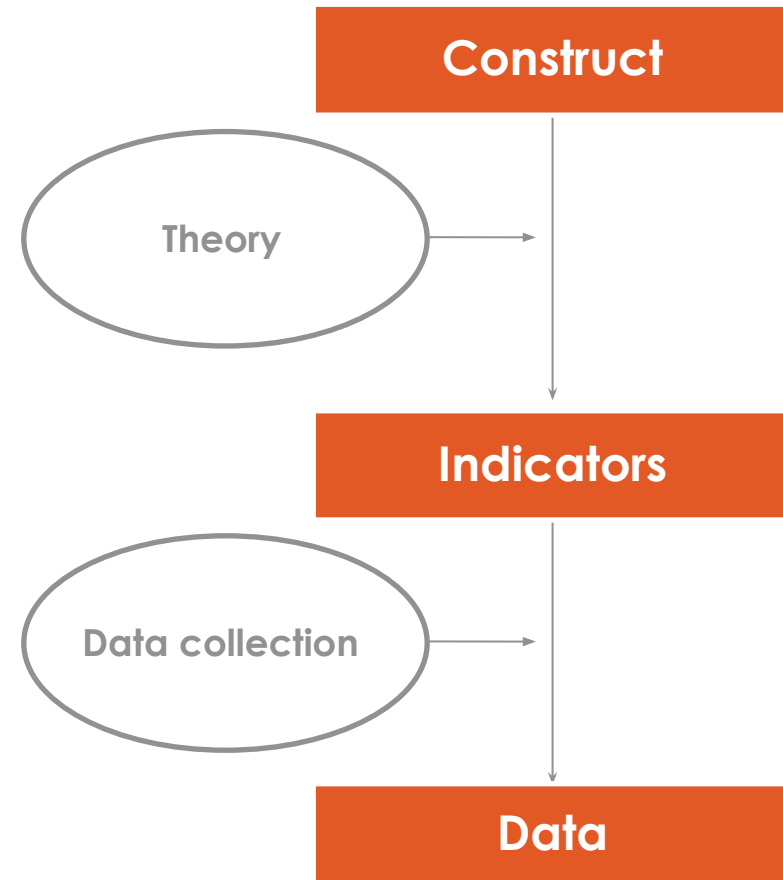


# Agenda

- I. ToC and stakeholder engagement
- II. Measurement recap and deeper dive
  - Conceptual principles of measurement
  - Sources of data
  - Best practices for data collection and use
  - SYEP example
- I. Process evaluation
- II. Q&A



# Concepts of measurement



The main concept being investigated. A construct is often abstract. (E.g., earnings, voice & representation, safety and security).

How you actually measure or “operationalize” your construct. (E.g., salary provides living wage).

What we use to measure our indicators.

# Menti

*Go to [menti.com](https://menti.com), enter code 5769 4418*



# Crime involvement is:

- A. A construct
- B. An indicator
- C. Data
- D. Don't know

*Go to [menti.com](https://www.menti.com), enter code 5769 4418*

# Violent crime arrest rate is:

- A. A construct
- B. An indicator
- C. Data
- D. Don't know

*Go to [menti.com](https://www.menti.com), enter code 5769 4418*

# Arrest records are:

- A. A construct
- B. An indicator
- C. Data
- D. Don't know

*Go to [menti.com](https://www.menti.com), enter code 5769 4418*

# 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 crime involvement, we could measure:
  - Total number of arrests
  - Or separate them by type:
    - Violent crime arrest rates
    - Property arrests rates
    - Drug-related arrests
- Two criteria to consider
  - **Validity**
  - **Reliability**

# Measurement criteria

- **Validity** – measuring the right thing

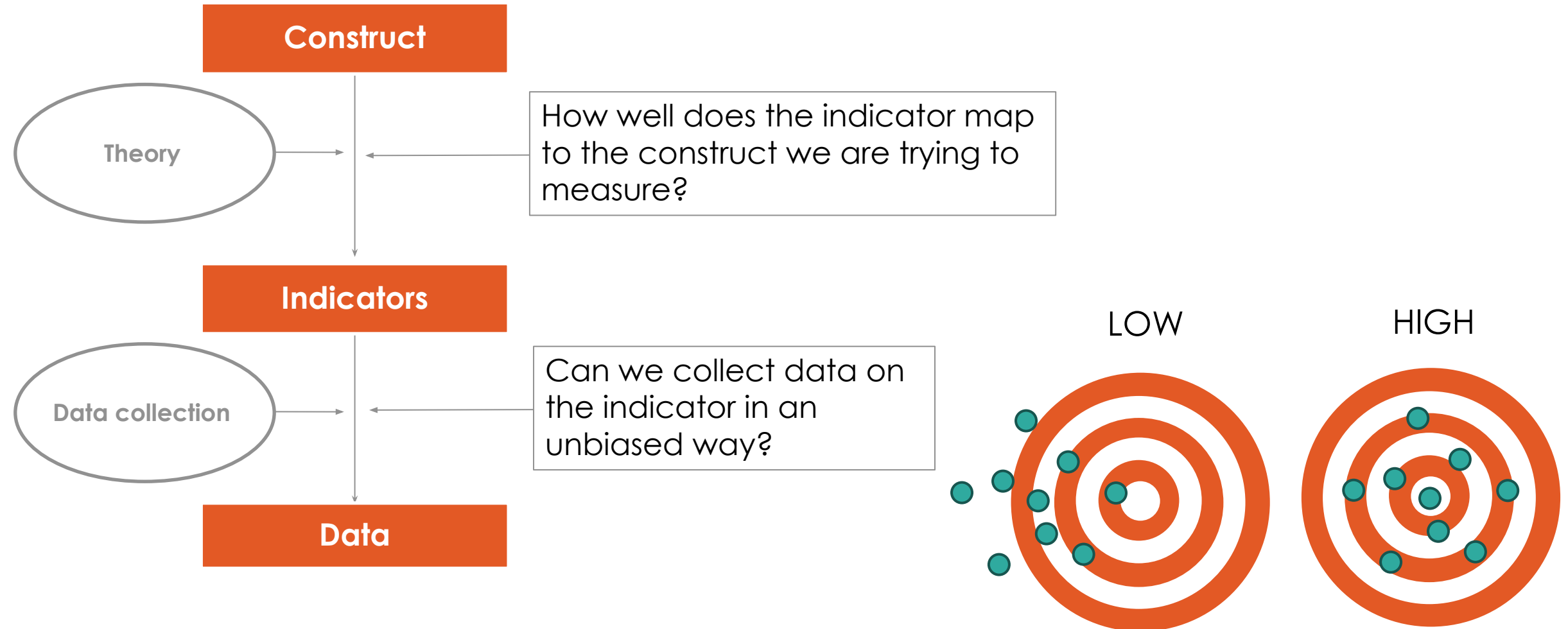


- **Reliability** – measuring the thing precisely





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

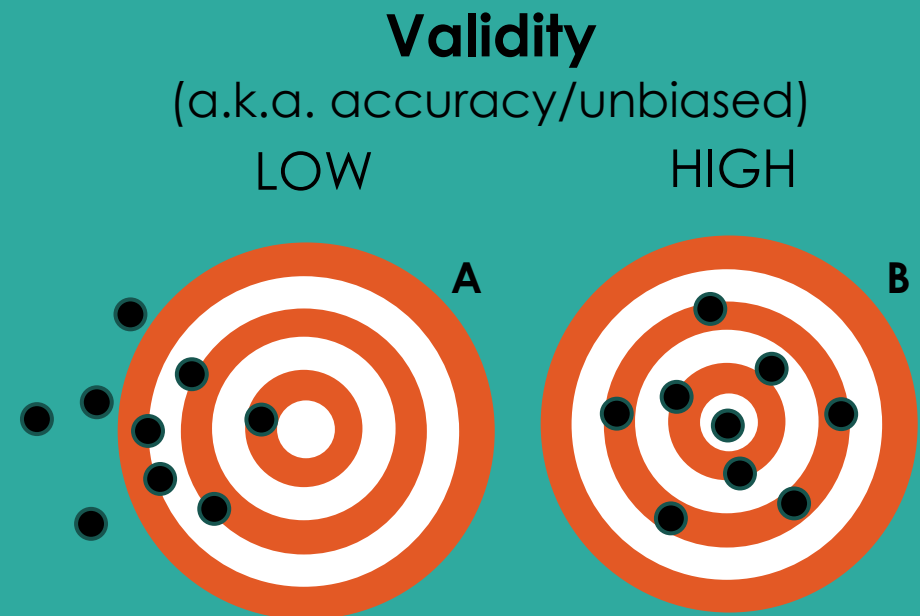


## Question:

Construct: Crime involvement

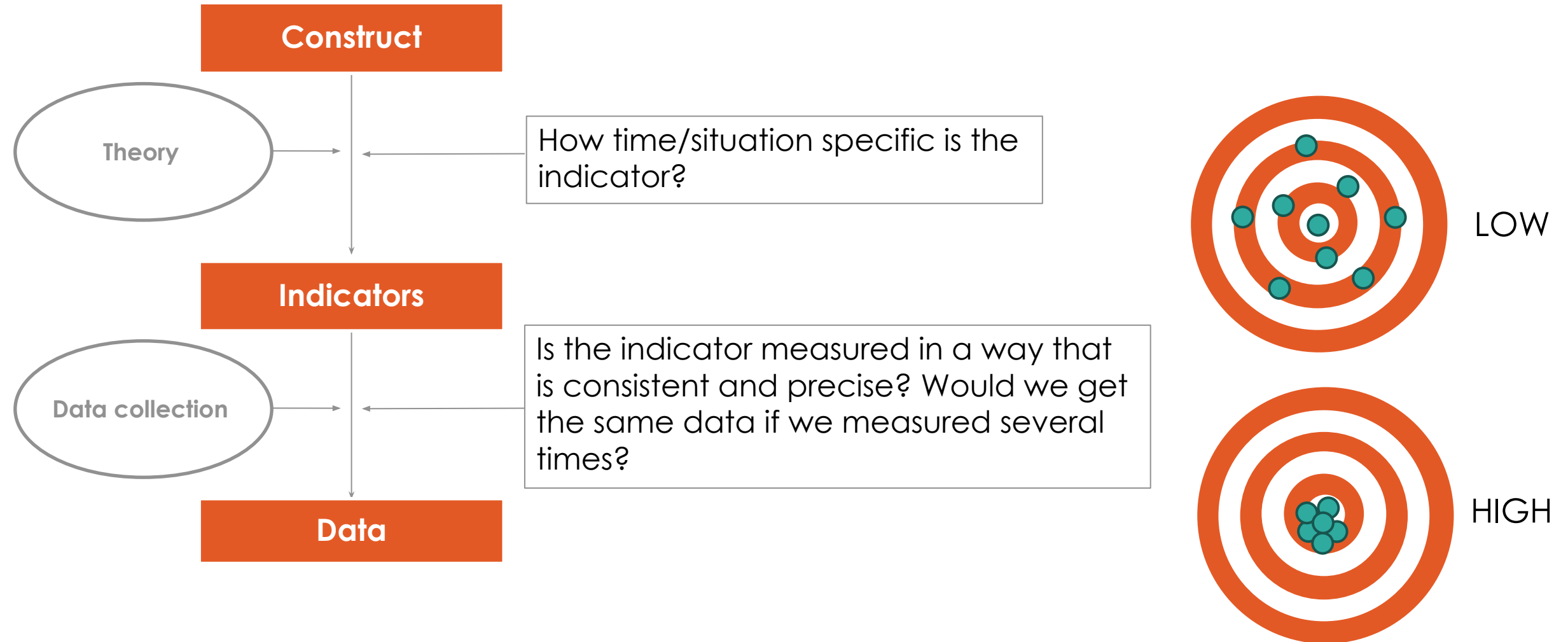
Indicator: Crime arrest rates

Where does that indicator land?



Go to [menti.com](https://www.menti.com), enter code 5769 4418

# Reliability (a.k.a. precision)



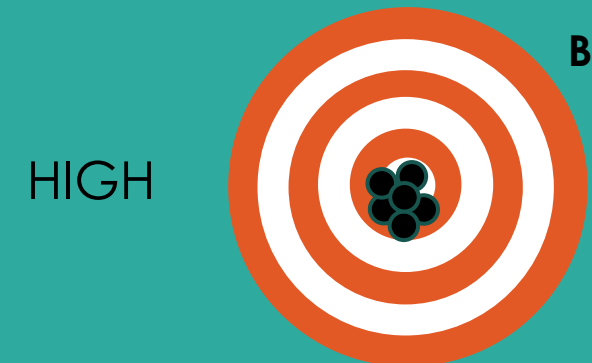
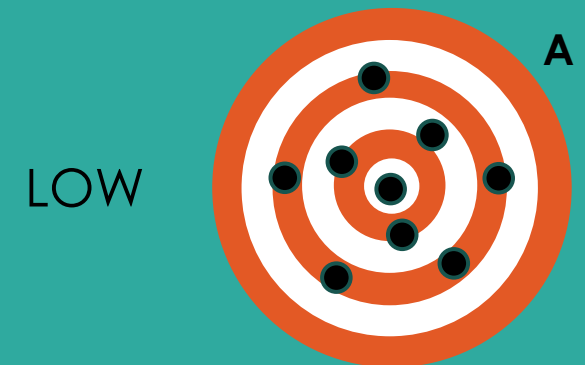
## Question:

Construct: Crime involvement

Indicator: Crime arrest rates

Where does that indicator land?

**Reliability**  
(a.k.a. precision)

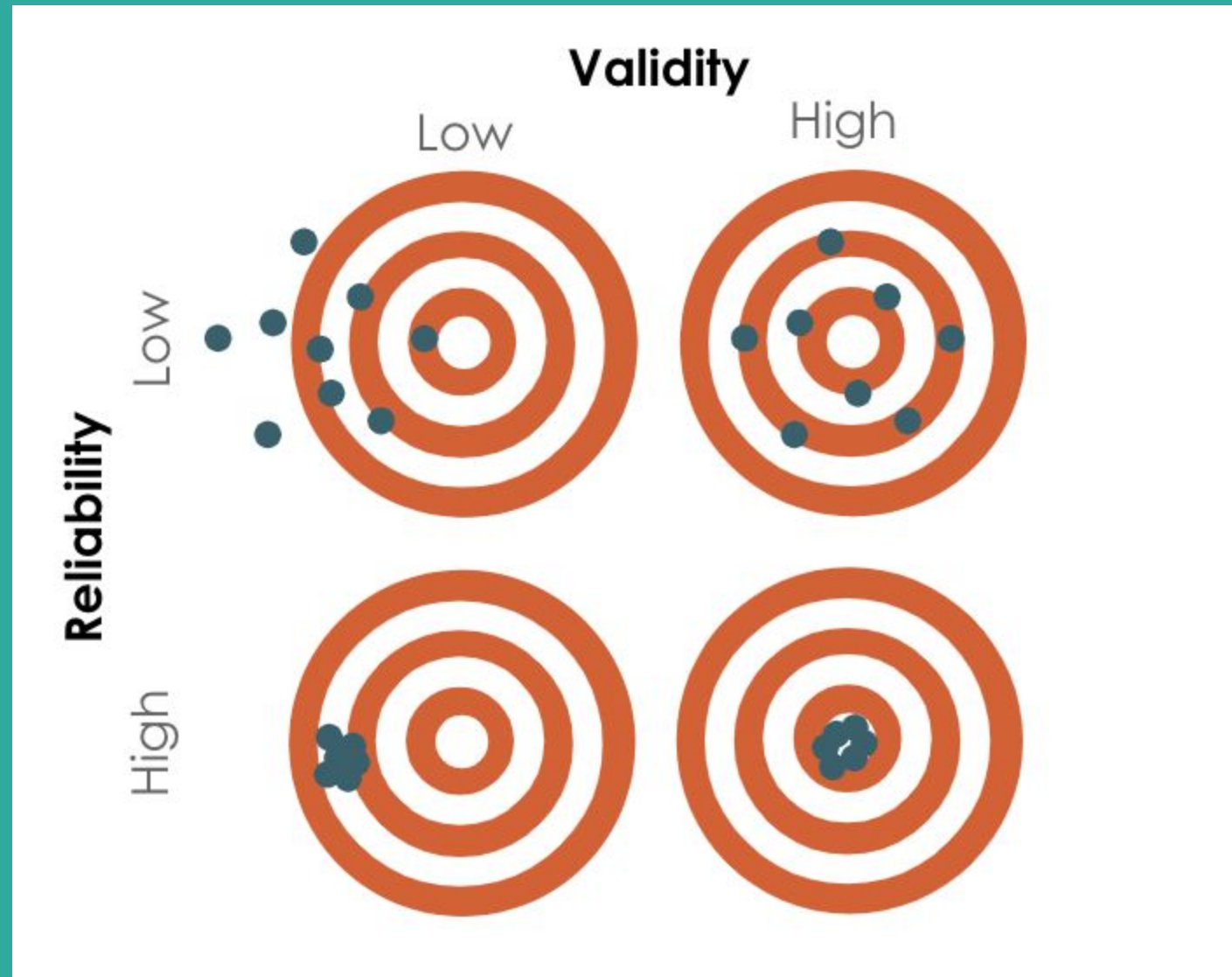


Go to [menti.com](https://www.menti.com), enter code 5769 4418

## 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



Go to [menti.com](https://www.menti.com), enter code 5769 4418

Before we move on:

**Are there any questions?**

# Agenda

- I. ToC and stakeholder engagement
- II. Measurement recap and deeper dive
  - Conceptual principles of measurement
  - Sources of data
  - Best practices for data collection and use
  - SYEP example
- I. Process evaluation
- II. Q&A

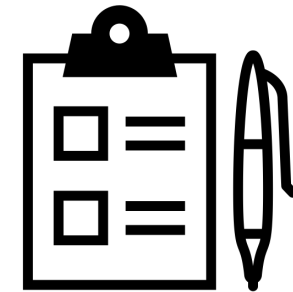
# Where can we get data?

## **Administrative Data** (from existing sources)

- Already exists; may exist retrospectively and prospectively
  - Enables long-term follow-up
- May include near census of relevant population
- Low logistical burden on researchers
- Lower burden on research participants
- Relatively inexpensive (usually)
- May reduce bias and error

## **Primary Data** (collected for a study)

- You have control: can tailor questions to your outcomes of interest
- Can ask about opinions, beliefs, and other subjective questions
- May be able to reach populations untracked in administrative data
- May reduce bias and error





# What are administrative data?

**Information collected, used, and stored primarily for administrative purposes (i.e., operational), rather than for research purposes. For example:**

- Unemployment insurance records
- IRS data
- WIOA
- Personnel records



# Why and how to use administrative data

## **Certain metrics may already be tracked in administrative data**

- Available retrospectively & prospectively
- Collected at time of occurrence
- Non-self-reported / passively collected
- Near census of relevant population
- Requires some up-front logistical work to secure Data Use Agreements (DUAs), digitizing data, and establish terms for data sharing

# Data use agreements (DUAs)

Documents the terms under which a data provider shares data with another entity

Most universities, as well as some implementing organizations and data providers, have their own templates

## Elements

- Project description
- Users and analysts
- Data security procedures
- Data to be shared
- Timeframe
- Data destruction
- Publication review
- Data publication

Before we move on:

**Are there any questions?**

# Agenda

- I. ToC and stakeholder engagement
- II. Measurement recap and deeper dive
  - Conceptual principles of measurement
  - Sources of data
  - Best practices for data collection and use
  - SYEP example
- I. Process evaluation
- II. Q&A

# Best practices for data collection

- Use validated survey/instrument measures
- Pilot surveys and data collection processes
- Ask questions about your data
- Validate your data

# Overarching tip: Get an early start

## **Plan to get administrative data before beginning the evaluation**

- Allow the availability of data to inform the analysis plan

## **Plan appropriately for data lags**

- Some data are available on a one-year or more lag
- Allow time for the data provider to extract data and transfer

## **Take advantage of the “down time”**

- Prepare analysis plan (e.g., code and staffing) to be ready once data arrive

# 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



Before we move on:

**Are there any questions?**

# Agenda

- I. ToC and stakeholder engagement
- II. Measurement recap and deeper dive
  - Conceptual principles of measurement
  - Sources of data
  - Best practices for data collection and use
  - SYEP example
- I. Process evaluation
- II. Q&A

# Summer Youth Employment Programs (SYEP)



Photo: University of Chicago Crime Lab

Based on "Summer Jobs Reduce Violence among Disadvantaged Youth" by Sara B. Heller (2014), *Science*; Stopping a Bullet With a Summer Job, J-PAL; and J-PAL Voices: The Impact and Promise of Summer Jobs in the United States, J-PAL.

# ToC and measurement of SYEP

**Needs**



**Inputs**



**Outputs**



**Intermediate Outcomes**



**Impact**

Theory of change	Measurement
Violent crime is a prevalent issue throughout the United States, with youth being twice as likely as adults to both commit and suffer from violent acts. Many have connected this violence to joblessness.	<b>Indicators:</b>  <b>Data:</b>
Youth are invited to apply to summer jobs programs.	<b>Indicators:</b>  <b>Data:</b>
Youth participate in the SYEP over the summer and receive social and emotional learning programming, and a steady paycheck.	<b>Indicators:</b>  <b>Data:</b>
Youth build social and emotional skills, and form social connections. Youth experience long-lasting changes in behavior and skills.	<b>Indicators:</b>  <b>Data:</b>
Reduction in criminal outcomes and/or dangerous behavior.	<b>Indicators:</b>  <b>Data:</b>



# ToC and measurement of SYEP

Needs



Inputs



Outputs



Intermediate Outcomes



Impact

Theory of change	Measurement
Violent crime is a prevalent issue throughout the United States, with youth being twice as likely as adults to both commit and suffer from violent acts. Many have connected this violence to joblessness.	<p><b>Indicators:</b> rates of crime and demographic information</p> <p><b>Data:</b> administrative records from the Chicago Police Department</p>
Youth are invited to apply to summer jobs programs.	<p><b>Indicators:</b> numbers of mentors, hours of SEL training implemented, employers</p> <p><b>Data:</b> program records about mentors, jobs, etc.</p>
Youth participate in the SYEP over the summer and receive social and emotional learning programming, and a steady paycheck.	<p><b>Indicators:</b> number of applications; attendance and participation</p> <p><b>Data:</b> admin records on applications; attendance sheets</p>
Youth build social and emotional skills, and form social connections. Youth experience long-lasting changes in behavior and skills.	<p><b>Indicators:</b> strength of peer network; connections with responsible/supportive adults (e.g., mentor, employer)</p> <p><b>Data:</b> interviews, surveys</p>
Reduction in criminal outcomes and/or dangerous behavior.	<p><b>Indicators:</b> arrest records spanning a year after the program ends</p> <p><b>Data:</b> individual-level infraction, or administrative records about arrests from the Chicago Police Department</p>

Before we move on:

**Are there any questions?**

# Agenda

- I. ToC and stakeholder engagement
- II. Measurement recap and deeper dive
- III. Process evaluation
- IV. Q&A

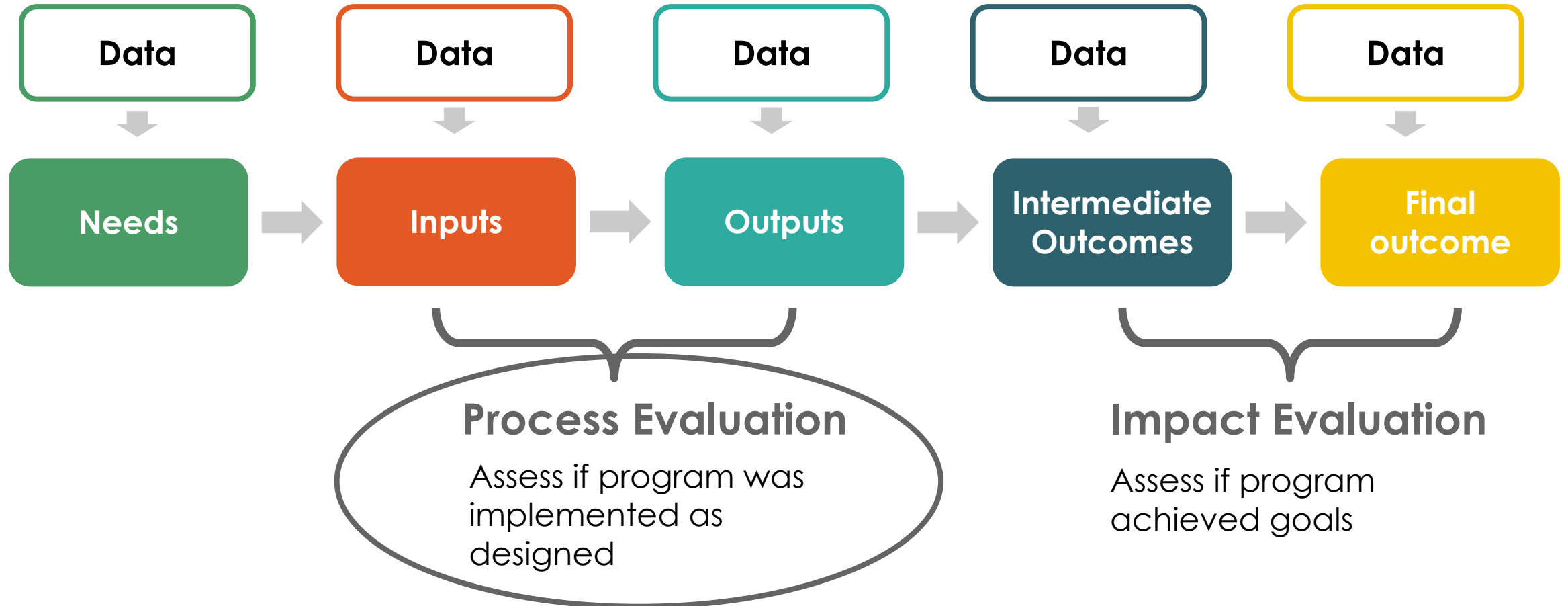
# Agenda

- I. ToC and stakeholder engagement
- II. Measurement recap and deeper dive
- III. Process evaluation
  - Key principles
  - Mapping an example
- I. Q&A





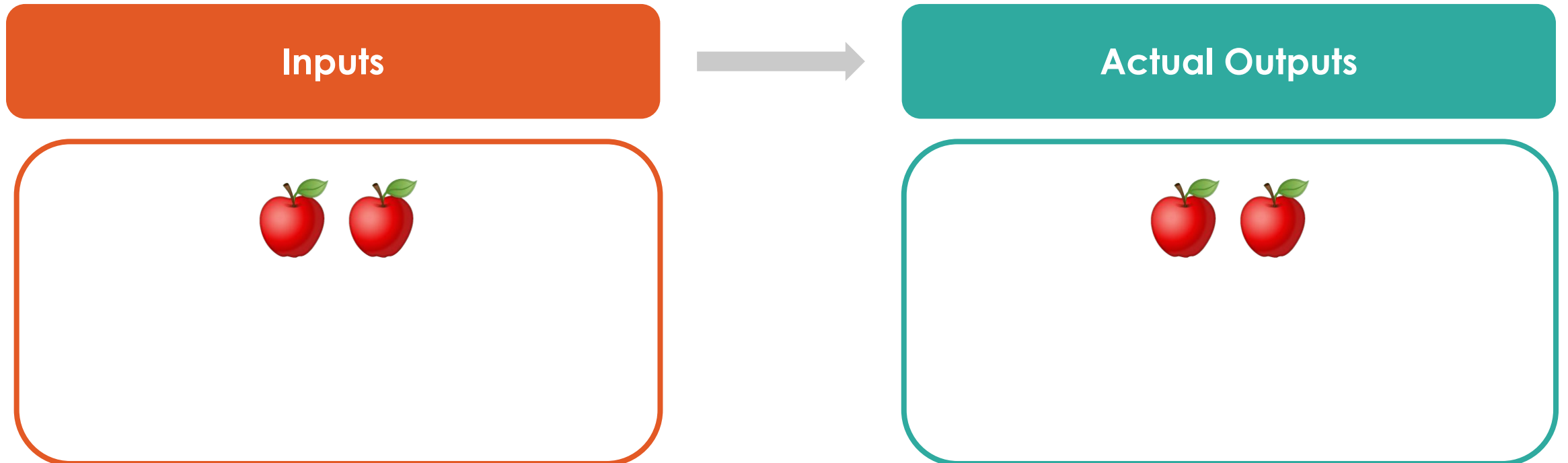
# A good impact evaluation builds on good program design and implementation



# Compare inputs & goals to outputs

Are you **meeting your goals** and objectives for **implementation**?

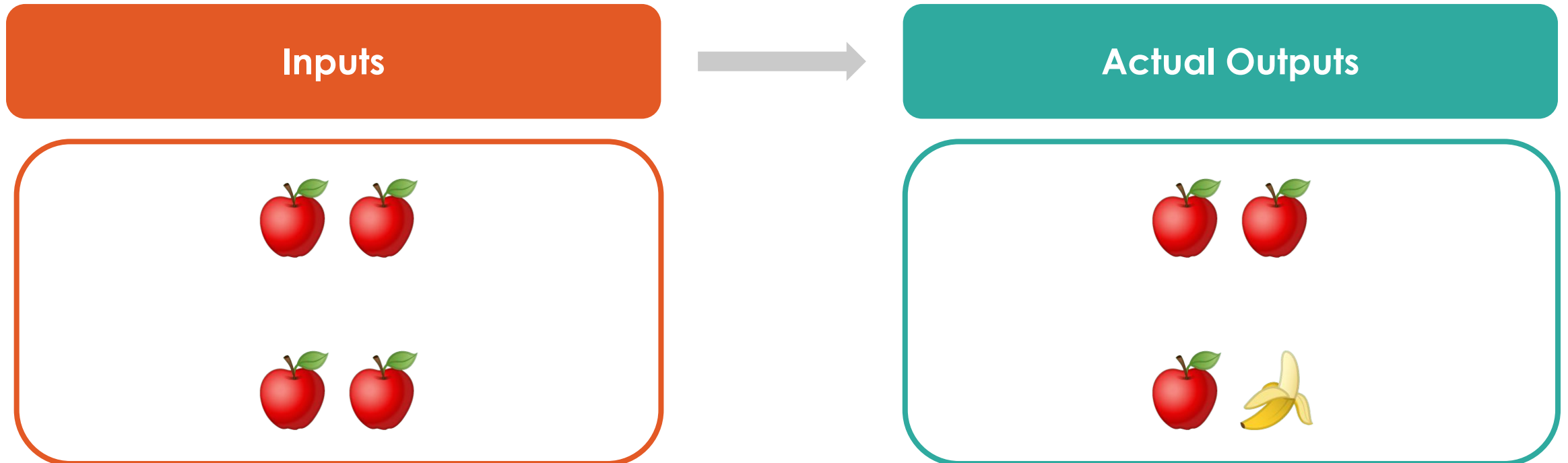
Are all of the elements of your program (inputs) actually being implemented?



# Compare inputs & goals to outputs

Are you **meeting your goals** and objectives for **implementation**?

Are all of the elements of your program (inputs) actually being implemented?



# Process evaluations

**What:** Process evaluations assess the extent to which a program is being **implemented with fidelity** to its design

In other words, are all of the elements of the program actually happening as intended?

**Why:**

**How:**

# Process evaluations

**What:** Process evaluations assess the extent to which a program is being **implemented with fidelity** to its design

In other words, are all of the elements of the program actually happening as intended?

**Why:** It's important to **identify any gaps** in implementation, inputs/components that are unrealistic in practice, and areas where the **program can be strengthened**.

If/when you go on to evaluate the *impact* of a program, it's helpful to know that your outcomes are a result of the program design and that this **design can then be replicated**.

**How:**

# How: key monitoring tools



Additional survey questions



Administrative data



Focus groups and qualitative interviews



High-frequency monitoring



Other objective, observable data



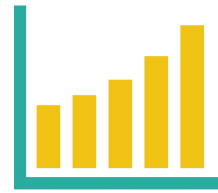
Site visits

# Five types of monitoring data



## Financial indicators

Understand how resources are allocated along the ToC. Supports analysis of cost effectiveness



## Activity tracking

Data on key activities and outputs from the ToC



## Targeting

Information on the people in a program to: i) identify who enters the program and ii), for those in the program, what service they should be provided.



## Take-up and engagement

Information about if people are actually using a product or service



## Feedback

Information about strengths and weaknesses of the program from the perspective of those it seeks to help

# CART principles

## Monitoring data should be:

- **Credible**

*It should be of high quality and believable*

- **Actionable**

*Commit to act on the data you collect*

- **Responsible**

*Ensure the benefits of data collection outweigh the costs. Administrative data that is already collected can help on this.*

- **Transportable**

*Collect data that will generate knowledge for other programs.*



# Integrate measurement into program design

Data collection can occur **regularly** (e.g., attendance) or **occasionally** (e.g., a site observation).

Proactively building data collection strategies into the design of a program itself can make the evaluation process easier.

Some suggestions:

- Include the collection process as part of the intervention protocols. For example, short “exit tickets” where participants share one thing they learned can double as both a content check and an attendance record
- Plan for any audits, site visits/observations, and other checks proactively
- Utilize any existing data

Before we move on:

**Are there any questions?**

# Agenda

- I. ToC and stakeholder engagement
- II. Measurement recap and deeper dive
- III. **Process evaluation**
  - Key principles
  - **Mapping an example**
- I. Q&A



# ToC and measurement of SYEP

Needs



Inputs



Outputs



Intermediate Outcomes



Impact

Theory of change	Measurement
Violent crime is a prevalent issue throughout the United States, with youth being twice as likely as adults to both commit and suffer from violent acts. Many have connected this violence to joblessness.	<p><b>Indicators:</b> rates of crime and demographic information</p> <p><b>Data:</b> administrative records from the Chicago Police Department</p>
Youth are invited to apply to summer jobs programs.	<p><b>Indicators:</b> numbers of mentors, hours of SEL training implemented, employers</p> <p><b>Data:</b> program records about mentors, jobs, etc.</p>
Youth participate in the SYEP over the summer and receive social and emotional learning programming, and a steady paycheck.	<p><b>Indicators:</b> number of applications; attendance and participation</p> <p><b>Data:</b> admin records on applications; attendance sheets</p>
Youth build social and emotional skills, and form social connections. Youth experience long-lasting changes in behavior and skills.	<p><b>Indicators:</b> strength of peer network; connections with responsible/supportive adults (e.g., mentor, employer)</p> <p><b>Data:</b> interviews, surveys</p>
Reduction in criminal outcomes and/or dangerous behavior.	<p><b>Indicators:</b> arrest records spanning a year after the program ends</p> <p><b>Data:</b> individual-level infraction, or administrative records about arrests from the Chicago Police Department</p>

# SYEP example

## Inputs

What are the key elements of the program?

- Delivery of SEL training: 10 hours x 7 weeks
- Summer job placement

# SYEP example

## Inputs

What are the key elements of the program?

- Delivery of SEL training: 10 hours x 7 weeks  
**Indicator:** Hours of SEL training implemented  
**Data:** Session rosters with start and end hours
- Summer job placement  
**Indicator:** Number of available jobs  
**Data:** List of jobs

# SYEP example

## Outputs

**Outputs** refer to what the participants are *actually* receiving as part of the program.

A few guiding questions:

- How many SEL sessions did the participants actually receive?
- How many participants were placed in jobs?
- How are you documenting these metrics?
- How often are you assessing these metrics? Are you able to make adjustments based on the data?

Is the  
intervention  
delivered as  
designed?

# SYEP example

## Outputs

Is the intervention delivered as designed?

- Delivery of SEL training: 10 hours x 7 weeks  
**Indicator:** Number of sessions attended  
**Data:** Session attendance lists
- Summer job placement  
**Indicator:** Number of participants placed on a job  
**Data:** Matchmaking records linking jobs to participants payment



# SYEP example

## Outputs

Is the intervention delivered as designed?

- Delivery of SEL training: 10 hours x 7 weeks  
**Indicator:** Number of sessions attended  
**Data:** Session attendance lists
- Summer job placement  
**Indicator:** Number of participants placed on a job  
**Data:** Matchmaking records linking jobs to participants payment

90%\* finished 7 weeks of the program

28 participants\* were placed in jobs, but no employment records were available for these youth

→ Overall: 75%\* of the youth who were offered the program actually participated

Before we move on:

**Are there any questions?**

# Adjust and improve as needed

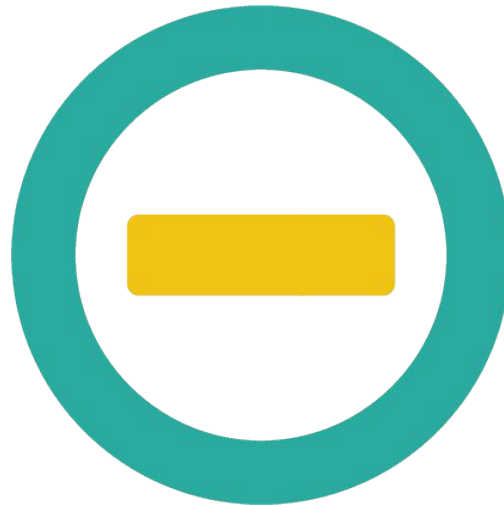
Sometimes things don't work out in practice the way we thought they would.

Use your process evaluation to determine:

**What went right**



**What went wrong  
(and why!)**



**How to make  
improvements**



Before we move on:

**Are there any questions?**

# Agenda

- I. ToC and stakeholder engagement
- II. Measurement recap and deeper dive
- III. Process evaluation
- IV. Q&A



Thank you!



# Acknowledgements

This case study is based on “Summer Jobs Reduce Violence among Disadvantaged Youth” by Sara B. Heller (2014), *Science*; *Stopping a Bullet With a Summer Job*, J-PAL; *J-PAL Voices: The Impact and Promise of Summer Jobs in the United States*, J-PAL; and *The Promises of Summer Youth Employment Programs: Lessons from Randomized Evaluations*, J-PAL.

J-PAL would like to thank the authors who have allowed us to use their papers, policy insights, and episodes as teaching tools.

## Further reading and resources

- J-PAL's [Introduction to Randomized Evaluations](#) Research Resource
- J-PAL's [Teaching resources on randomized evaluations](#)
- IPA's [Goldilocks Toolkit on Theory of Change: Laying the Foundation for Right-Fit Data Collection](#)
- J-PAL's [Measurement & Data Collection](#) Research Resource
- J-PAL's [Survey Design](#) Research Resource
- J-PAL's [Repository of Measurement and Survey Design Resources](#)



## Further reading and resources

- J-PAL's [Communicating with a Partner About Results](#) Research Resource
- J-PAL's [Implementation Monitoring](#) Research Resource

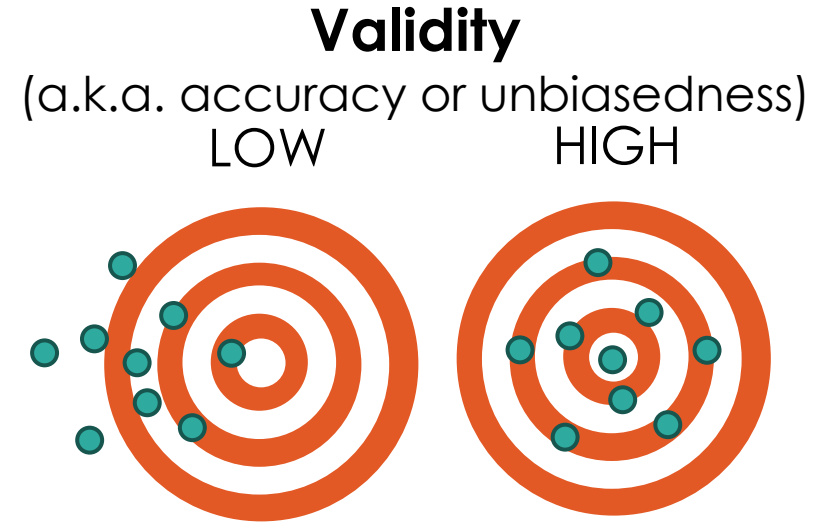
# Appendix



# 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)

**Reliability**  
(a.k.a. precision)

LOW



HIGH



## 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](#)