Development
Learning
Lab

Measurements

Catalina Franco, SNF & FAIR at NHH

DLL Incubator, December 6, 2022

CMI CHR. MICHELSEN INSTITUTE
UNIVERSITY OF BERGEN
NHH
SNF

Our Roadmap

- → **Why** measure?
- → **What** to measure?
- → **Where** to find the data?
- → **When** to collect the data?
- → **How** to collect the data?

Why measure

- Allows to assess:

1. Was there animpact of the program? Was it positive or negative? What is thesize of the effect?
2. Why there was an impact or no impact

- Otherwise: going "blind" without knowing whether the program works of why it works

What, Where, When and How to Measure
Case study: Wheels of Change program

Motivation and Background

Despite considerable progress in closing the gender gap in education, there still exists several **barriers to human capital accumulation for girls in developing countries**. Some of the important barriers include:

- cost of schooling,
- distance to school,
- safety,
- lack of agency,
- deep-rooted cultural norms.

In **rural Zambia,** three times as many girls in Grade 11 left school compared with their male peers.
In our sample, 98% of the students walk to school, taking 110 minutes each way, and 35% of the girls had been harassed on their commute.

What is the research question?

Does providing bicycles to girls who live far from their school affect **empowerment** and **educational attainment** ?

Main assumption:

The bike directly alleviates the costs of education: distance and safety

Wheels of Change: Transforming Girls' Lives with Bicycles
Nathan Fiala, Ana Garcia Hernadez, Kritika Karula and Nishith Prakash (2020)

Program

Bicycle for Empowerment and Education program (BEEP)

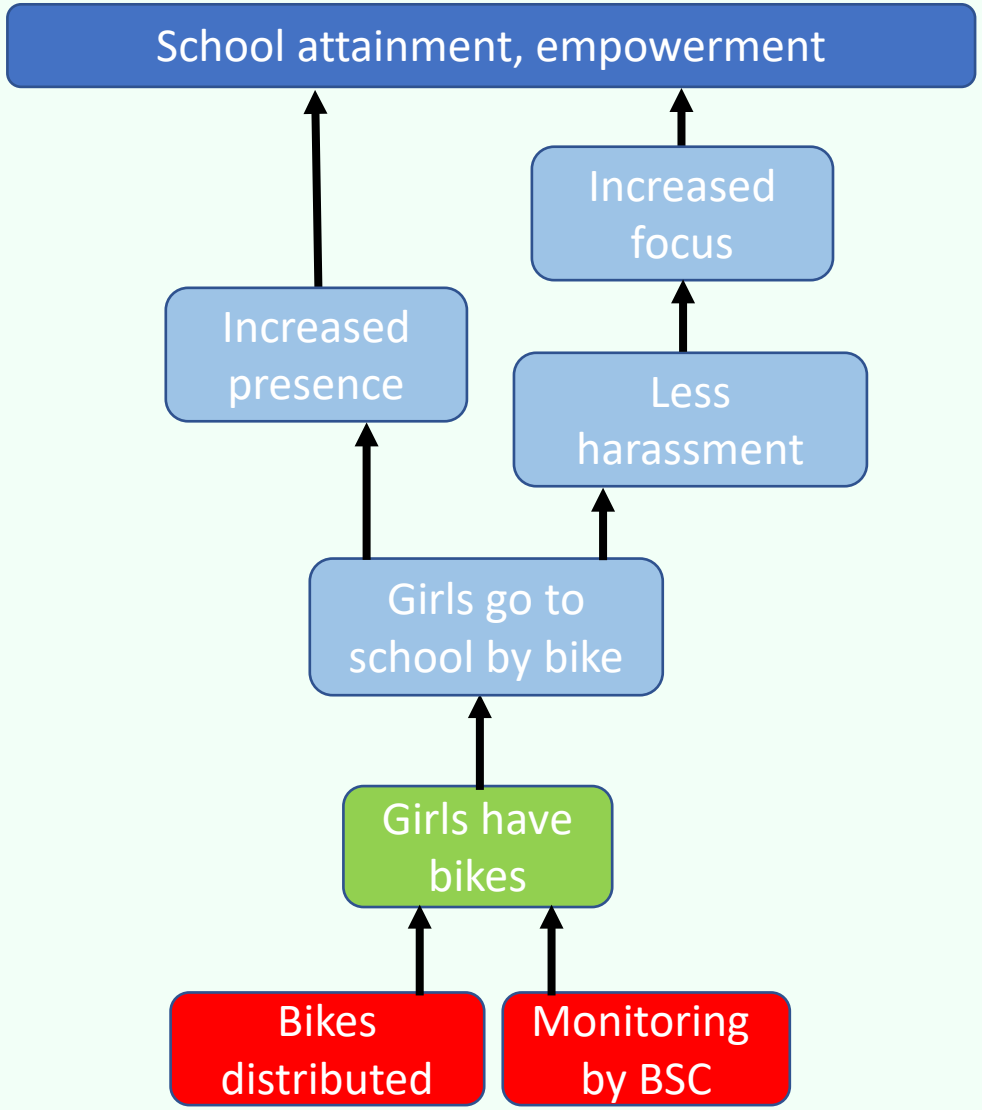
Bicycles to adolescent girls in grades 5,6, and 7
Live more than 2.5 km away from school
Must be used primarily to go to school (sign a contract)
Bicycle Supervisory Committee to monitor the program
Field mechanic trained for each school
Parents pay for reparations/maintenance

Impact Evaluation Design

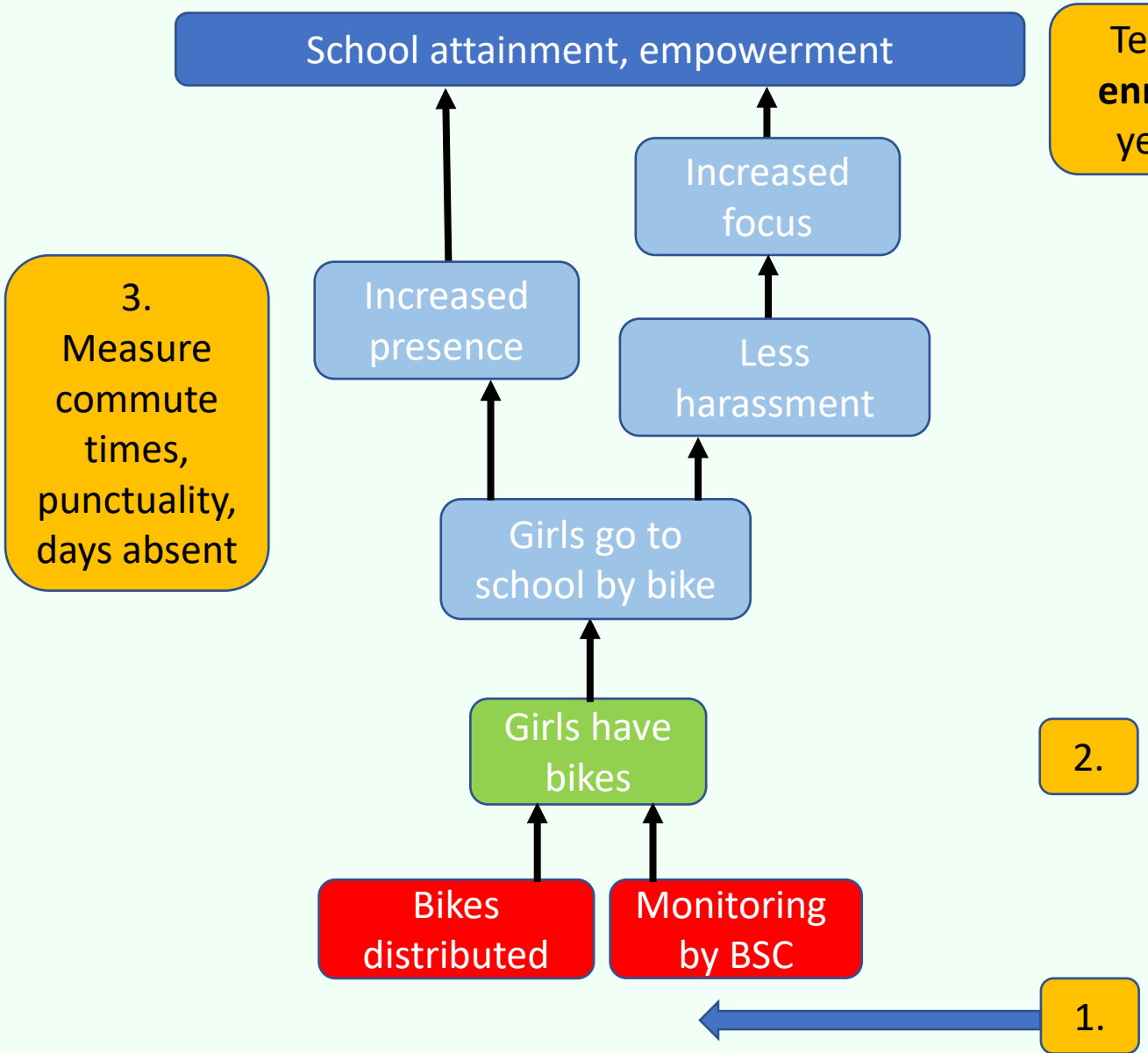
RCT in 100 schools in the Southern province of Zambia

T1 (25 schools)	T2 (20 schools)	Control (55 schools)
BEEP + upfront payment by parents	BEEP but no payment	Girls did not receive bicycles

Theory of change: Wheels of Change



What to measure: Wheels of Change



Test scores, whether **enrolled in school** 2-3 years after program

Empowerment index: control of their own decisions, self-image

In an RCT, if randomization is done properly, the only **required data** is the **endline** outcome measures

However, you may want to:

1. **Target** beneficiaries (e.g., girls who live atleast 2.5 km away)
2. Know whether **randomization worked** : Are people in the treatment and control groups similar in baseline characteristics (age, gender, etc.)
3. Know if the **changes/assumptions** in your theory of change are realized
4. Know more about how the **implementation** works

Some "trade-offs" on measurement

- High cost of omitting an important variable
- Low cost of adding an additional question in your survey

How many people to survey with the same budget?

- Survey same participants at baseline and endline
- Survey twice as many participants at endline only

Ideally we know what you are going to measure **before** we start the program:

- Questions are required as part of IRB approval
- Pre-analysis plans
- Always pilot questions!

Where to find the data

What is your unit of analysis?

→ A person (student, adult, child, teacher, etc.)

→ A household

→ An institution (a school, a municipality, a committee, etc.)

→ A business

→ An area of land

Multiple units of analysis are possible

Unit of analysis ≠ level of randomization:
Student vs. School in Wheels for Change

Main sources of data

Primary data:

- Surveys

→ Focus groups, interviews

→ Observation

→ Games, lab-in-the field experiments

→ Audits

Secondary data:

→ Administrative data (e.g., schools, government)

→ Web scraping, satellite images

→ Surveys conducted by others (e.g., household surveys, DHS)

When to collect the data

Baseline surveys

- → Targeting
- → Increasing statistical power : collecting data on the same people before and after program increases correlation and reduces noise
- → Heterogeneous impact of a program
- → Assessing balance between treatment and control
- → Dealing with attrition (loss of participants)

Midline surveys

- → How impact is evolving over time
- → May be expensive

Endline surveys

- → Minimum to assess the impact of a program

Data in Wheels for Change

WHAT	WHEN	SOURCE
Distance to school	Baseline	Survey
Demographic characteristics	Baseline	Survey
School attendance	Midline	Administrative data from schools
Safety: being whistled at or teased	Endline	Survey
Score in academic tests	Endline	Administrative data from schools or can administer own test

Pool the data to deal with different cohorts

Year 1	Year 2	Year 3
Treatment group		
Control group		
	Treatment group	
	Control group	
		Treatment group
		Control group

Year 1	Year 2	Year 3
Treatment group	Treatment group	Treatment group
Control group	Treatment group	Treatment group
Control group	Control group	Treatment group
Control group	Control group	Control group

How to collect the data

Use already existing scales

- → Example: Big-5 personality tests
- → Validated in previous research (sometimes shorter versions)
- → Saves time
- Comparable to other research
- → Find them in previous research papers

Use your own indicators

Good indicators are SMART:

- → Specific: Narrow and describes what needs to be measured
- → Measurable: Measured in the same way regardless of who uses the indicator
- → Achievable: Straightforward to collect the data
- → Relevant: Closely linked to the relevant outcome
- → Time-bound: time frame of collection or measurement

Keep in mind when designing survey questions: Framing effects

- Framing the same situation in a positive or negative way can change the way participants respond

- Even the exact terms or the order of the terms can matter

- Example:

Version A: "Individuals are more to blame than social conditions for crime and lawlessness in this country"

Version B: "Social conditions are more to blame than individuals for crime and lawlessness in this country"

➡ 60 and 57% agreed with the statement, respectively

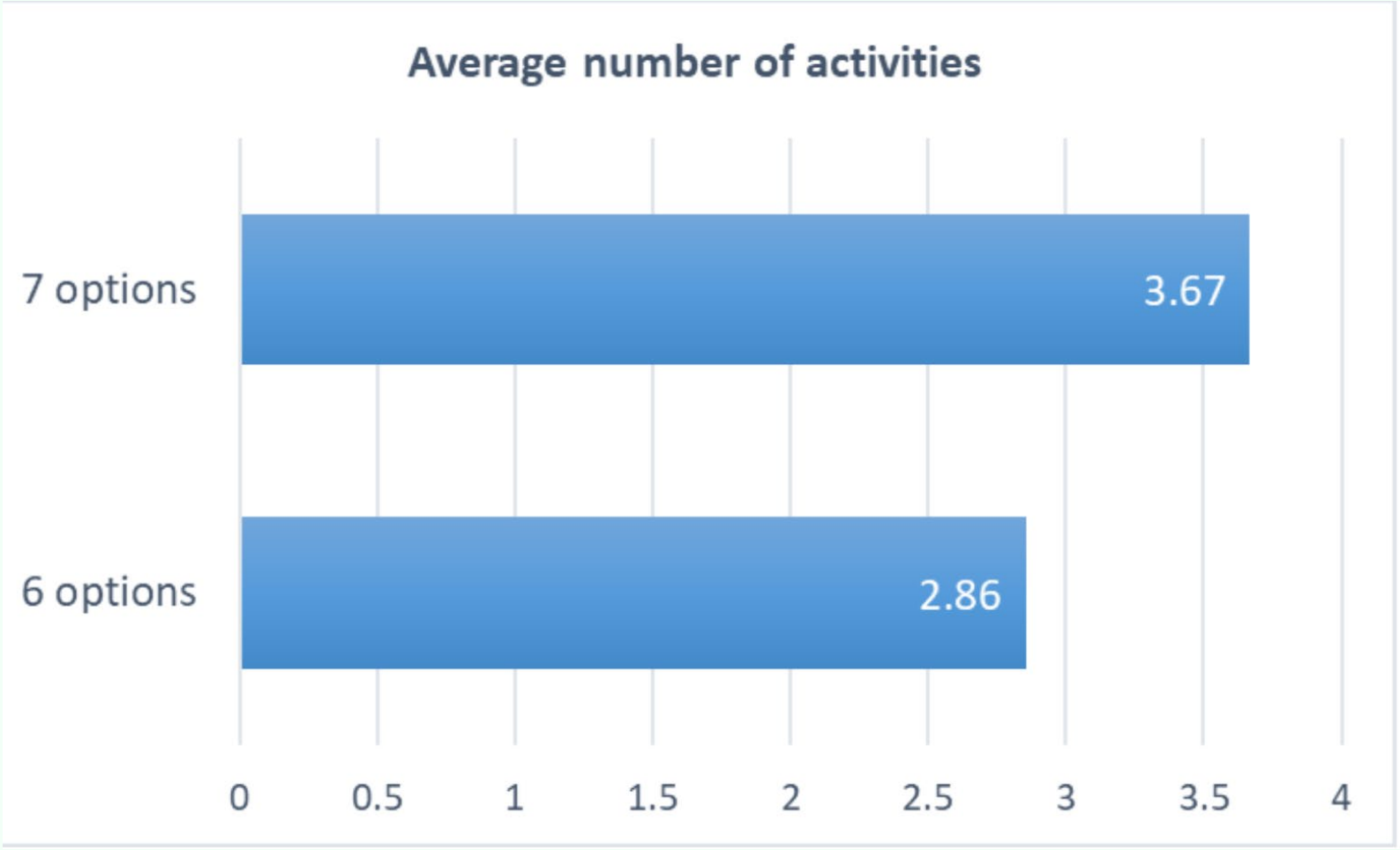
Keep in mind when designing survey questions: Social desirability bias

- People want to give a certain impression of themselves
- Appropriate or socially-accepted response even if untruthful
- Especially important with sensitive questions: we suspect a higher likelihood of answering untruthfully than other questions
- Randomize number of options in a list experiment

People often engage in risky behavior. How many of the following activities have you ever done?

1. Driving a car while drunk
2. One-night stand without a condom
3. Swimming/diving with sharks
4. Not respecting COVID rules (treatment group)
5. Smoking weed
6. Skydiving
7. Participating in a fight

Keep in mind when designing survey questions: Social desirability bias



81%

Our survey:

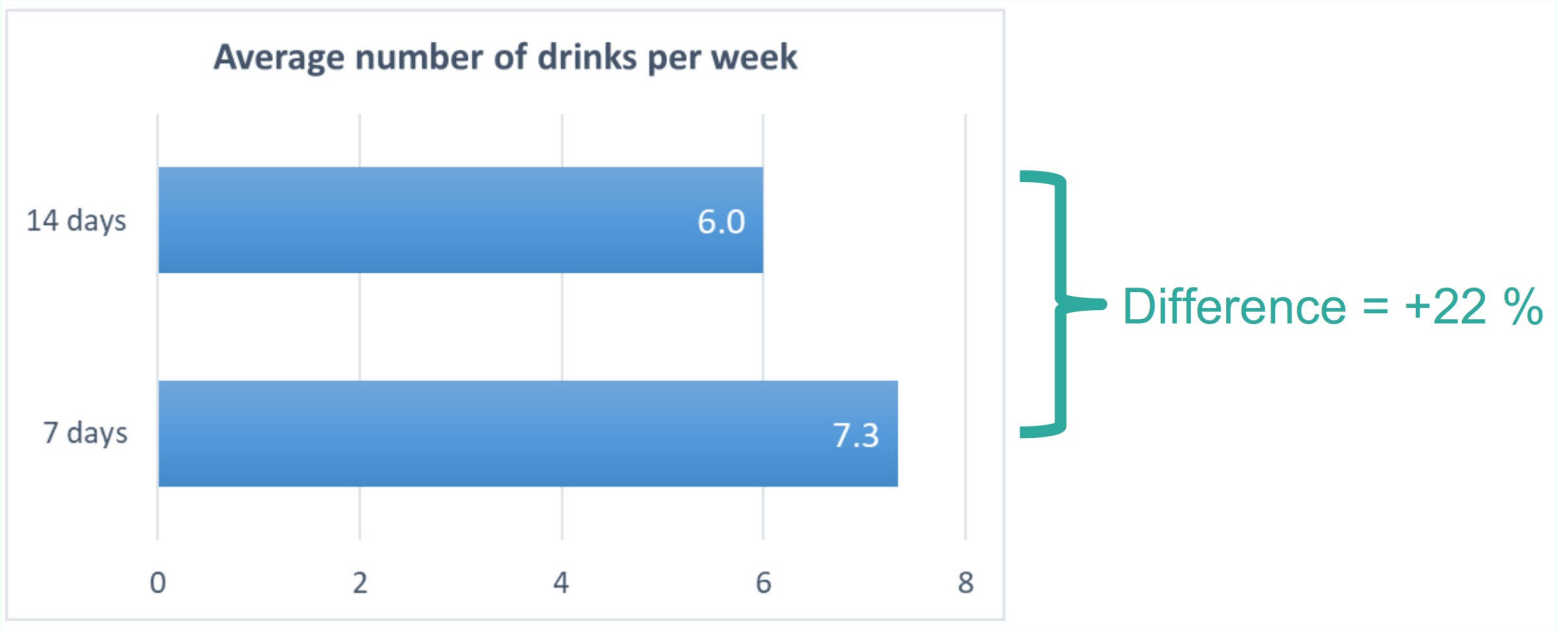
7 options:
2.4 (n=14)

6 options:
1.9 (n=12)

Keep in mind when designing survey questions: Recall bias

- Accuracy of recall may be influenced by factors such as the recall period
- Randomize the recall period

How many beers or glasses of wine did you drink during the last 7/ 14 days?



Our survey:

14 days: **4.5** (n=14)

7 days: **3.3** (n=12)

Thank you!

All the best with your measurements

Group work: Measurement

Recap: What is youmain outcomes of interest?

1. How will you measure it?

- If using an existing scale, what are the component of this scale?
- If creating your own indicator, is it SMART (Specific, Measurable, Achievable, Relevant, Time-bound)?

2. What type of data will you use to measure it? Describe in detail (e.g., admin. data from which source and collected when; if index, what components will it have, etc.)

Development
Learning
Lab

CMI CHR. MICHELSEN INSTITUTE

UNIVERSITY OF BERGEN

NHH

SNF