

Introduction to randomized evaluation

ดร.สุภาวรี บุญมานันท์
RCT workshop I





Outline

- What is program evaluation?
- Key steps in the RCT program evaluation
- Impact evaluation methods
- Randomized evaluation
- Measurement



Program evaluation


An assessment whether the program **achieves** its **goal**

Program	Immediate goal	Ultimate goal
Health: COVID-19 vaccination	Higher rate of vaccinated population	Lower rate of transmission, severe illnesses
Education: conditional cash transfer	Higher enrollment rate to higher education	Better employment outcomes

Source: Dr. Nada Wasi



Common approaches to program evaluation

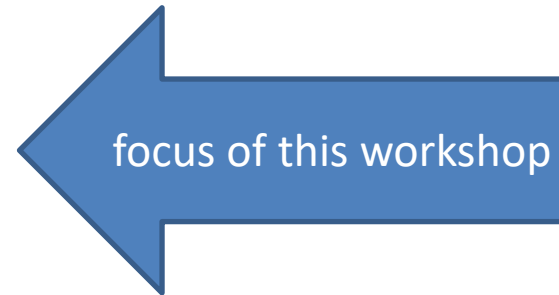
- Randomized Controlled Trial (RCT) 
- Observed data which mimic an experiment
- Model the mechanism, behavior and simulate the impact of intervention

Source: Dr. Nada Wasi



Common approaches to program evaluation

- Randomized Controlled Trial (RCT)



Cases possible for RCT when ...

- evaluation is planned alongside the intervention,
- process can be monitored,
- outcomes can be measured.

Source: Dr. Nada Wasi



Cases we cannot use RCT

- Unethical: intervention mentally or physically harmful to participants
- Programs had already happened, e.g. minimum wage jump in April 2012
- Outcome is very long run, e.g. pension reform
- Macro policies – nation-wide policy, e.g. increasing interest rate

Source: Dr. Nada Wasi

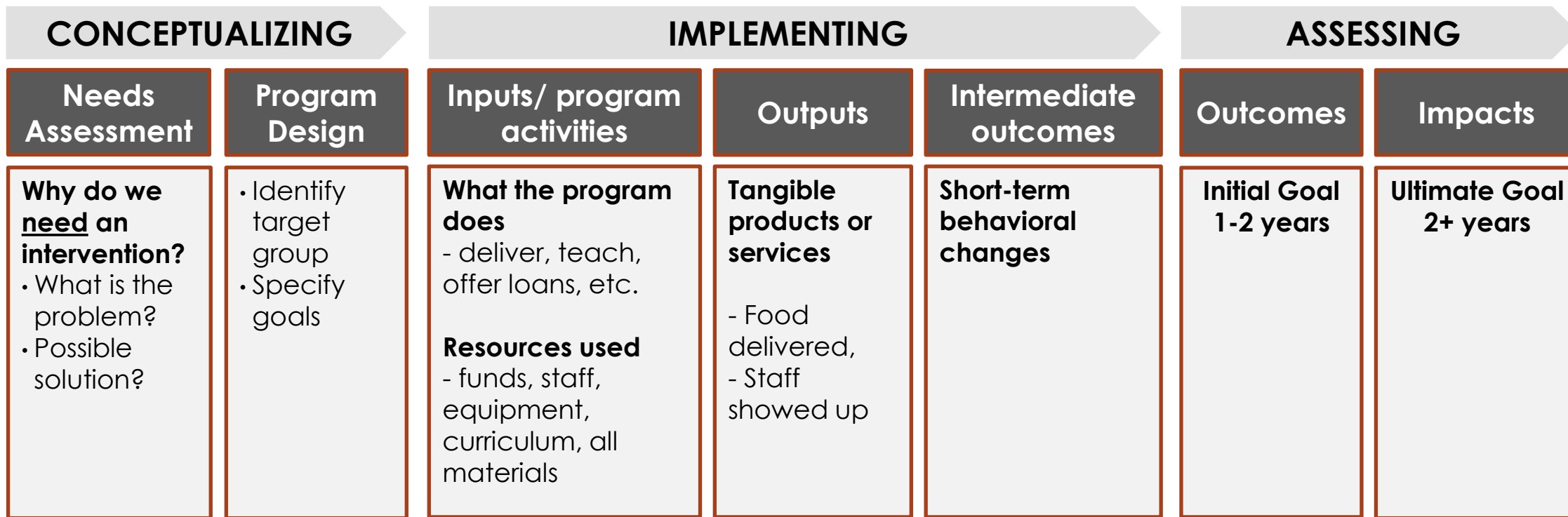


Outline

- What is program evaluation?
- **Key steps in the RCT program evaluation**
- Impact evaluation methods
- Randomized evaluation
- Measurement



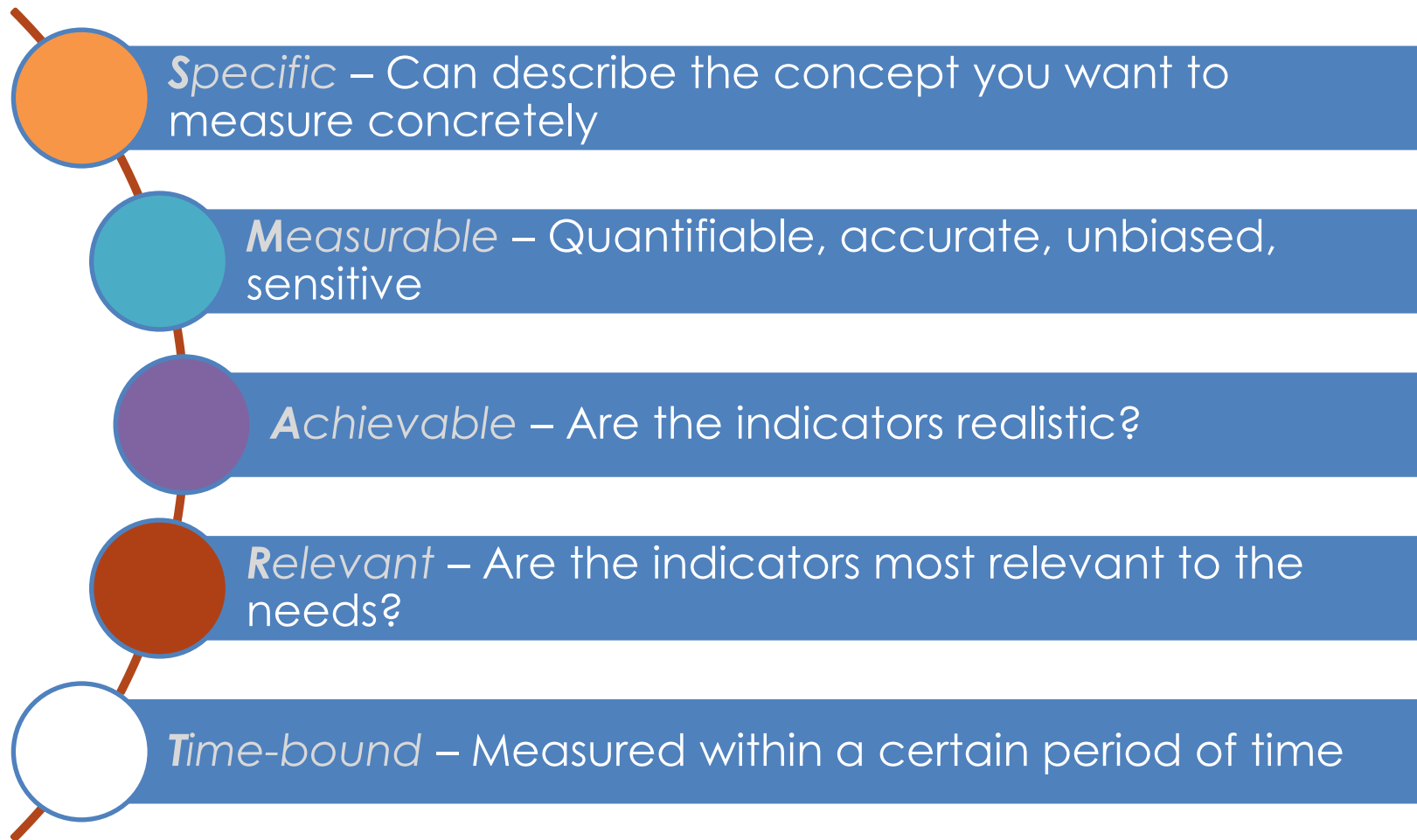
Key steps in the RCT program evaluation



Source: Abdul Jameel Poverty Action Lab



Identifying SMART Indicators



Source: Abdul Jameel Poverty Action Lab



Outline

- What is program evaluation?
- Key steps in the RCT program evaluation
- **Impact evaluation methods**
- Randomized evaluation
- Measurement



Basic vocabulary for impact evaluation

Causal inference / causality - ความเป็นเหตุเป็นผล

- ในการประเมินผลโครงการ “ความเป็นเหตุเป็นผล” มีความหมายต่างจากที่ใช้ในชีวิตประจำวันทั่วไป
- ในที่นี้ “ความเป็นเหตุเป็นผล” คือ
เมื่อเราสามารถพูดได้ว่า ความเปลี่ยนแปลงที่เราสังเกตเห็น ที่เราวัดได้ เกิดจากโครงการของเราเพียงอย่างเดียวเท่านั้น โดยไม่มีปัจจัยอื่นใด ๆ ที่ส่งผลต่อการเปลี่ยนแปลงนั้น



เราจะวัดผลของโครงการได้อย่างไร

ผลของโครงการ (impact) คือ ความแตกต่างระหว่าง

- สิ่งที่เกิดขึ้นจริง (what actually happened) และ
- สิ่งที่น่าจะเกิดขึ้น (what would have happened) ถ้าไม่มีโครงการนี้ หรือที่เรียกว่า **counterfactual**



Counterfactual

- **Counterfactual** คือ สถานการณ์ในโลกอีกใบที่ผู้เข้าร่วมโครงการจะได้ประสบ เมื่อในโลกนั้นไม่ได้มีโครงการ (หมายถึง เมื่อผู้เข้าร่วมโครงการไม่ได้เข้าโครงการนั่นเอง)
- **ปัญหา:** Counterfactual ไม่ได้เกิดขึ้นจริง เราจึงไม่สามารถเห็นมันได้
- **ทางออก:** เราจะต้อง “เลียนแบบ (mimic)” หรือสร้าง counterfactual ขึ้นมา ให้เป็นกลุ่มเปรียบเทียบ

การเลือกกลุ่มเปรียบเทียบ

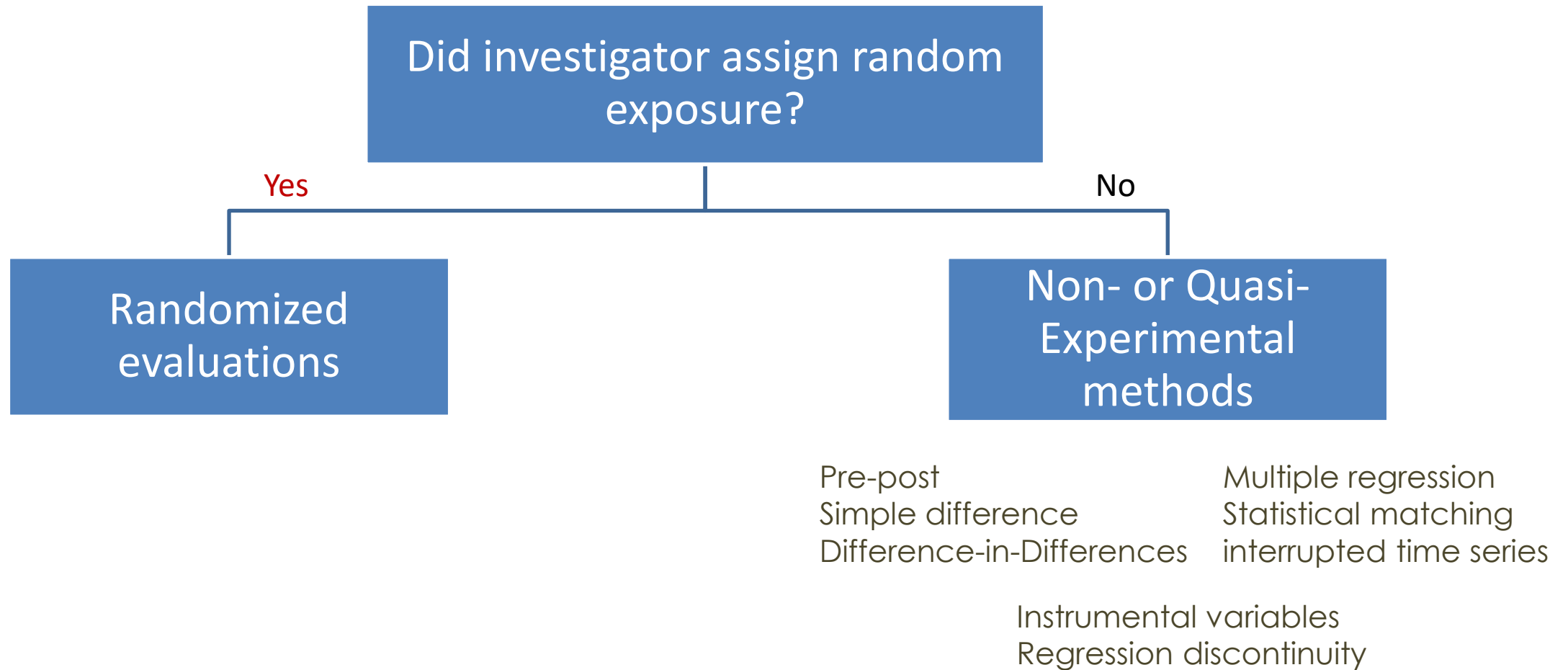
- **Idea:** เลือกกลุ่มเปรียบเทียบที่เหมือนกับกลุ่มผู้เข้าร่วมโครงการทุกประการ ยกเว้นเพียงอย่างเดียวคือกลุ่มเปรียบเทียบไม่ได้เข้าร่วมโครงการ



- **Goal:** เพื่อที่จะสามารถบอกได้ว่าความแตกต่างระหว่างกลุ่มใน outcome ที่เราสนใจ มาจากการได้เข้าโครงการ (และไม่ใช่มาจากปัจจัยอื่น)
- ดังนั้น คุณภาพการประเมินผลโครงการจึงขึ้นอยู่กับคุณภาพของการเลือกหรือการสร้างกลุ่มเปรียบเทียบให้เลียนแบบ **counterfactual** ได้สมจริง



Impact evaluation methods





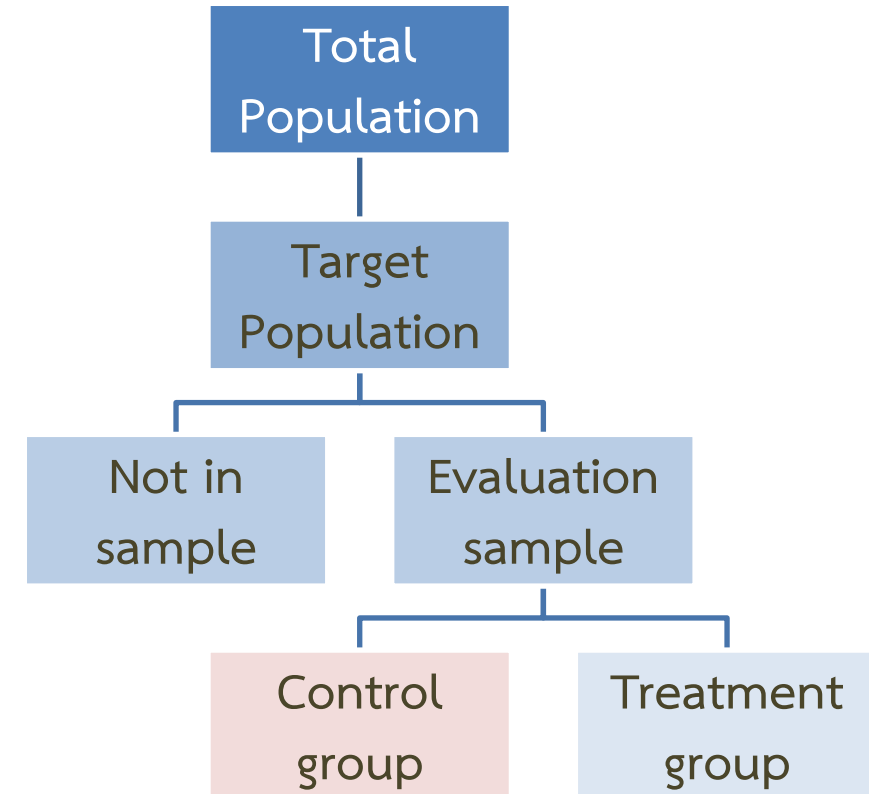
Outline

- What is program evaluation?
- Key steps in the RCT program evaluation
- Impact evaluation methods
- **Randomized evaluation**
- Measurement

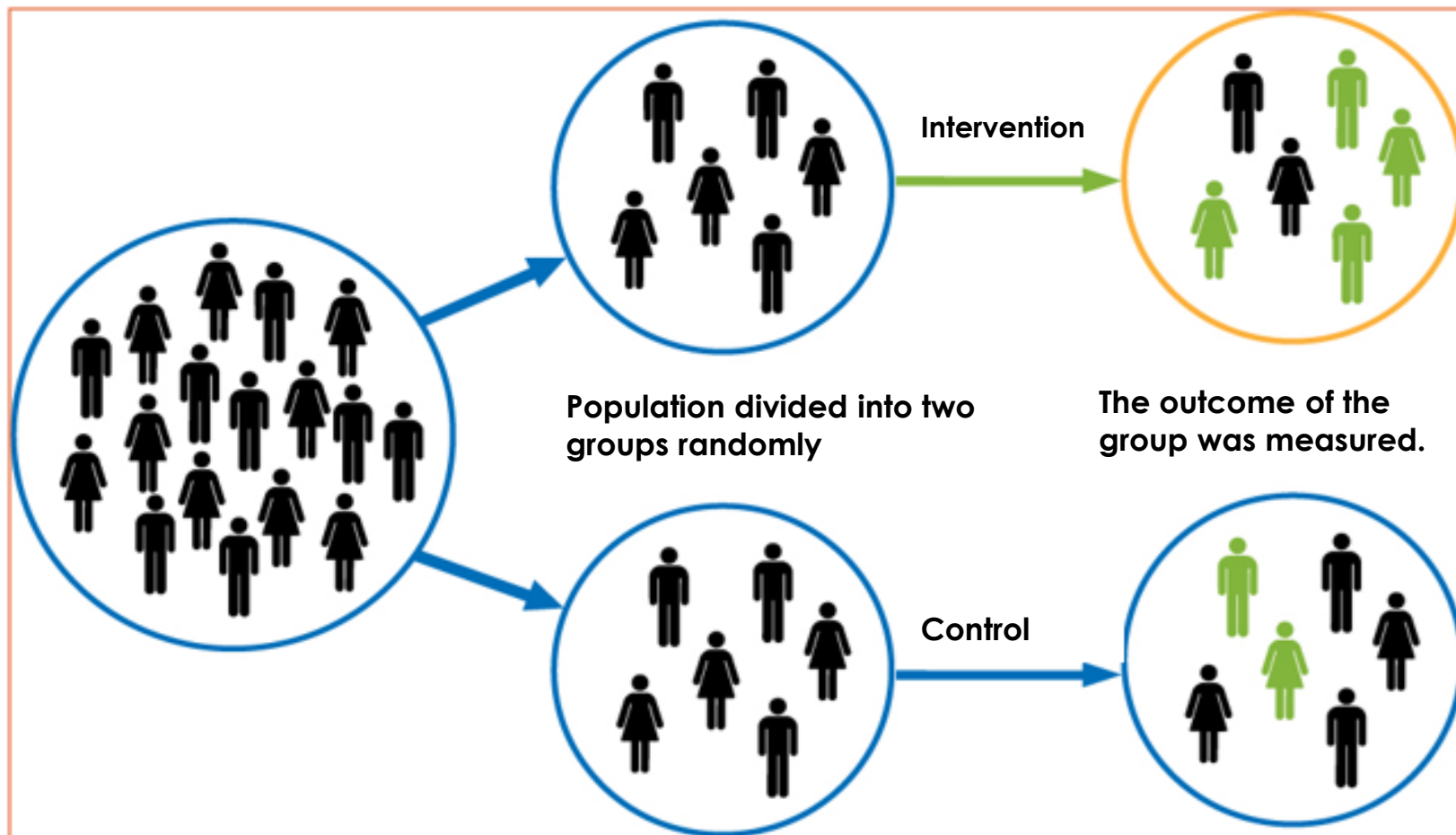


Randomized Evaluation – How to carry out?

- Suppose we evaluate the learning recovery program using randomized evaluation.
- Take a sample of program applicants.
- Then **randomly** assign them to either:
 - **Treatment group** receives the program.
 - **Control group** is not allowed to receive the program (during the evaluation period).



Randomized Evaluation – Why randomize?



If properly designed and conducted, randomized experiments can generate **two groups that do not differ systematically.**



=> provide a **credible** method to estimate the impact of a program



Randomized Evaluation - Importance of Sample Size

- Theoretically, random assignment should produce two groups that are equivalent or similar but with **large enough** sample size.
- **Law of large numbers:** As we increase the numbers of experiments, the observed value and the expected value converge.



Key advantage of randomized experiment

- **If properly designed and conducted**, two groups do not differ systematically at the outset of the experiment on characteristics that are observable (e.g., age, gender, income, test score) or unobservable (e.g., motivation, toughness, optimism).
- Any difference that subsequently arises between them can be **attributed** to the program rather than to other factors.
- With baseline data, we should always check as much as we can whether the randomization works.
 - **Balance test**: compare characteristics of participants in treatment and control groups before the program starts.





Outline

- What is program evaluation?
- Key steps in the RCT program evaluation
- Impact evaluation methods
- Randomized evaluation
- **Measurement**



การประเมินผลโครงการฟื้นฟูการเรียนรู้ของนักเรียนใน จ.สมุทรสาคร

- **ปัญหา:** นักเรียนใน จ.สมุทรสาคร ได้รับผลกระทบจากการปิดเรียนในช่วงการระบาดของโรคโควิด 19 จนทำให้เกิดการเรียนรู้ถดถอย โดย จ.สมุทรสาคร เป็นพื้นที่ที่มีการปิดเรียนยาวนานที่สุดในประเทศไทย
- **Intervention** โดยมูลนิธิโรงเรียนสตาร์ฟิชคันทรีโฮม
 - หลักสูตรอบรมเพื่อพัฒนาความรู้และความสามารถของครูในด้าน Blended Learning, Active learning, การใช้เทคโนโลยีสารสนเทศและการสื่อสารเพื่อสนับสนุนการเรียนรู้
 - สนับสนุนค่าใช้จ่ายสำหรับปัจจัยพื้นฐานของครูและนักเรียน และ internet package
 - สื่อดิจิทัลแบบ Micro Learning ซึ่งมีเนื้อหาส่งเสริมการอ่านเขียน คณิตศาสตร์ และการพัฒนาทักษะด้านอารมณ์และสังคม สำหรับครู ผู้ปกครองหรือตัวแทนชุมชน และผู้เรียนกลุ่มเป้าหมาย

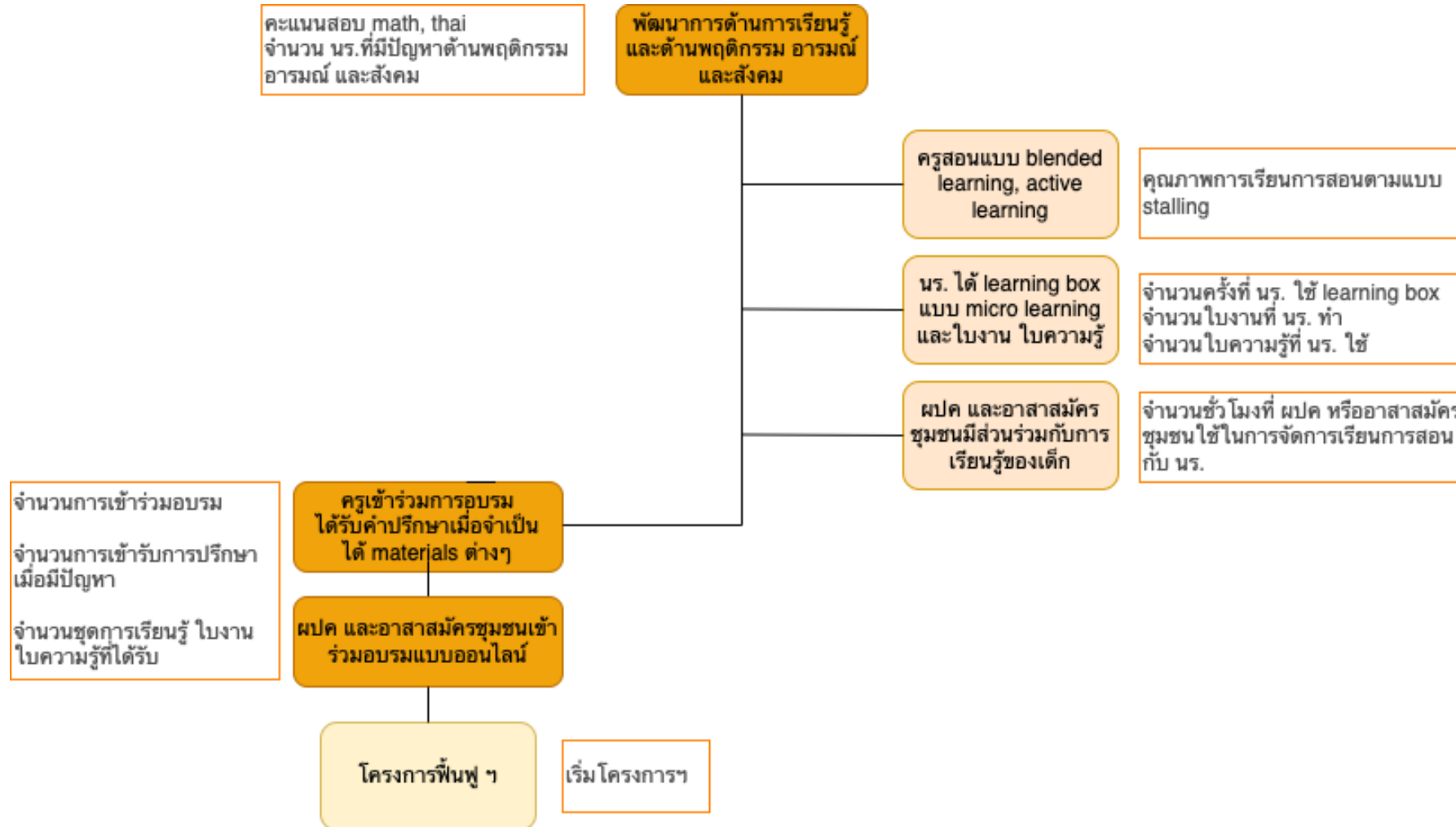


การประเมินผลโครงการฟื้นฟูการเรียนรู้ของนักเรียนใน จ.สมุทรสาคร

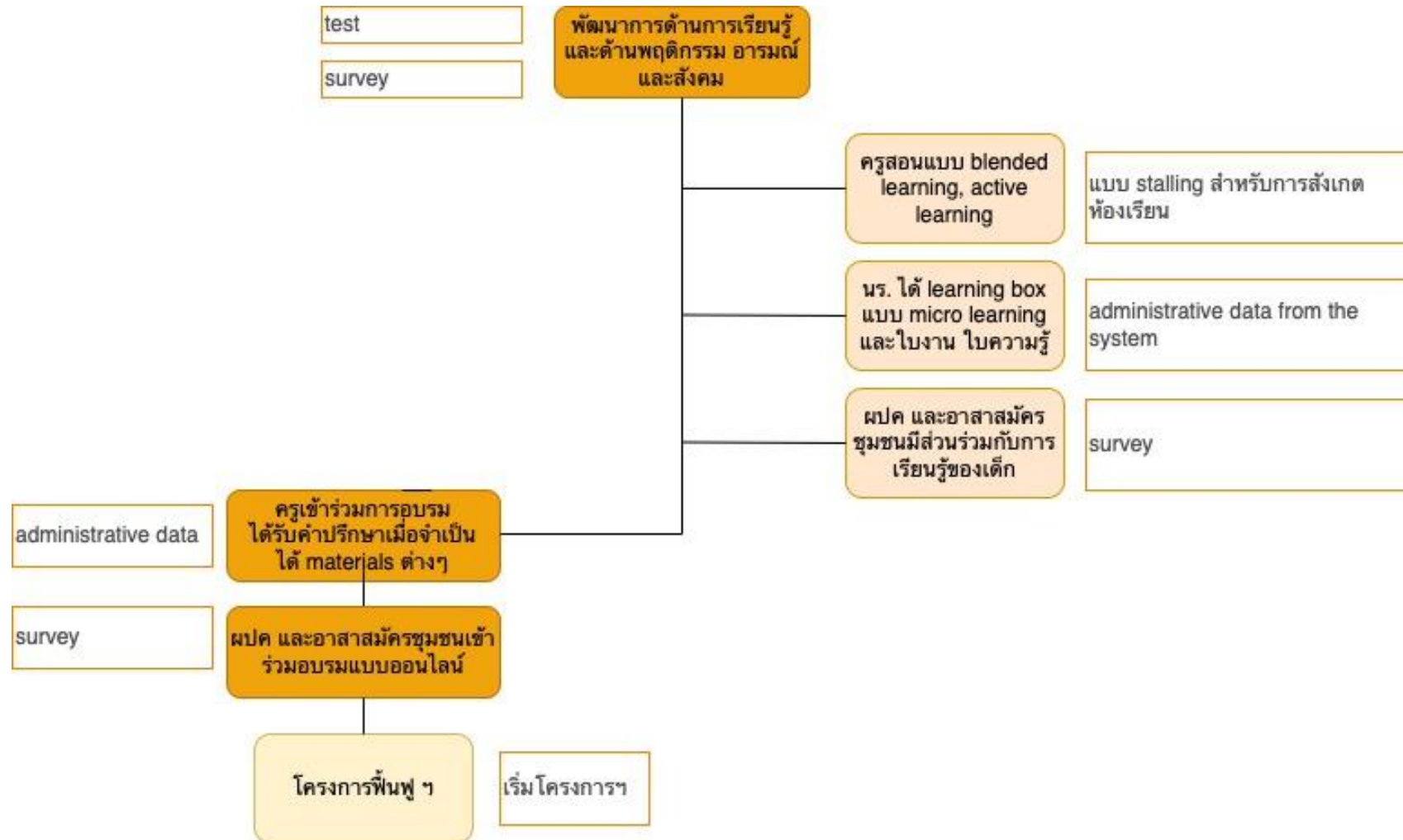
? โครงการฟื้นฟูการเรียนรู้ของนักเรียนใน จ.สมุทรสาคร มี impact อะไรต่อพัฒนาการด้านการเรียนรู้ของนักเรียน

- โครงการฯ ช่วยให้นักเรียนได้รับชุดการเรียนรู้ ใบงาน และใบความรู้มากขึ้นหรือไม่
- โครงการฯ ช่วยกระตุ้นให้ครูจัดการเรียนการสอนให้มีคุณภาพมากขึ้นหรือไม่

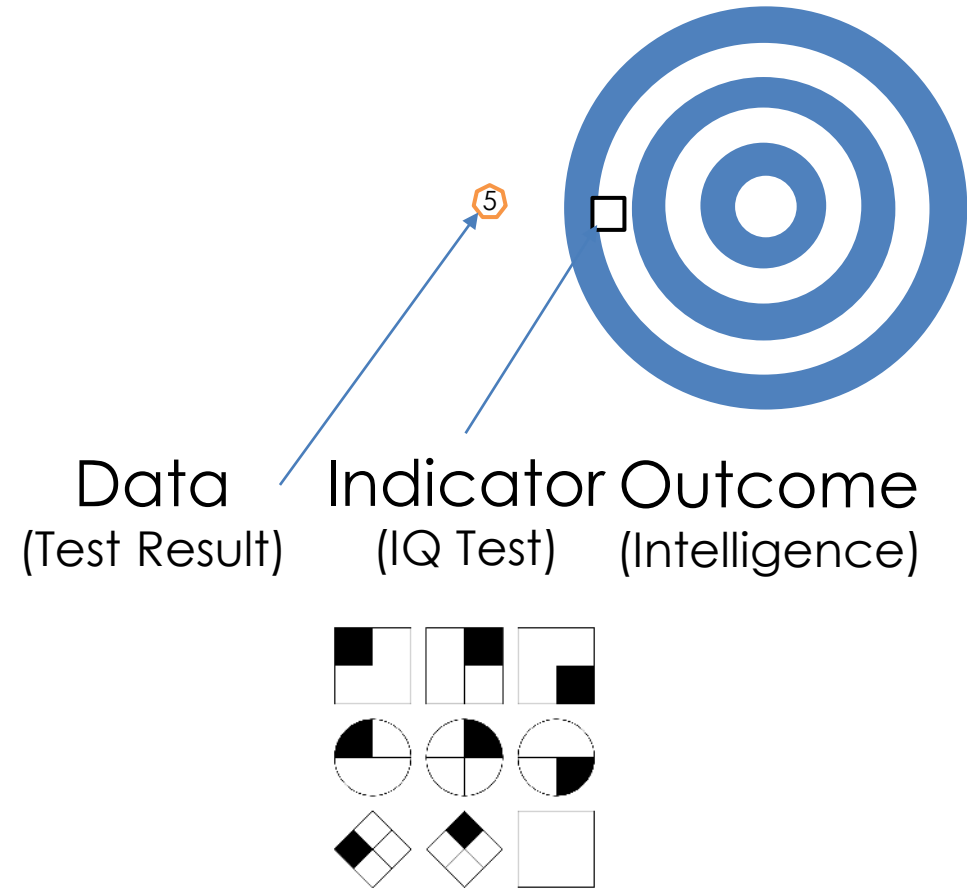
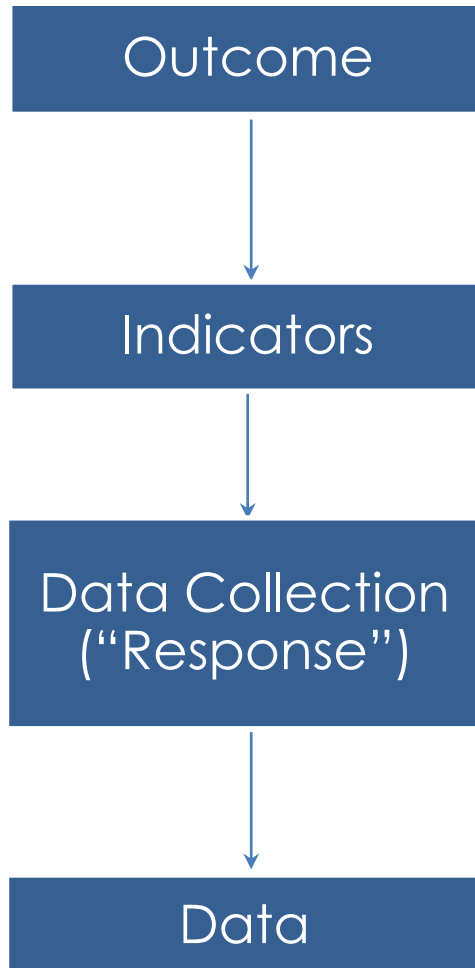
What to measure?



Data sources

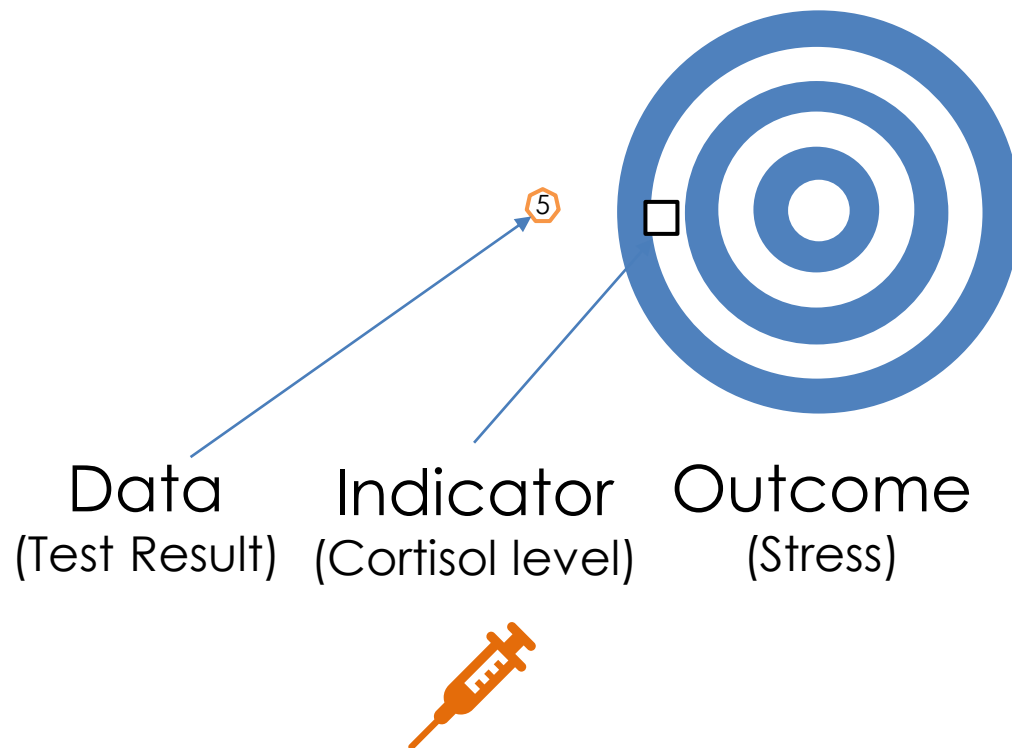
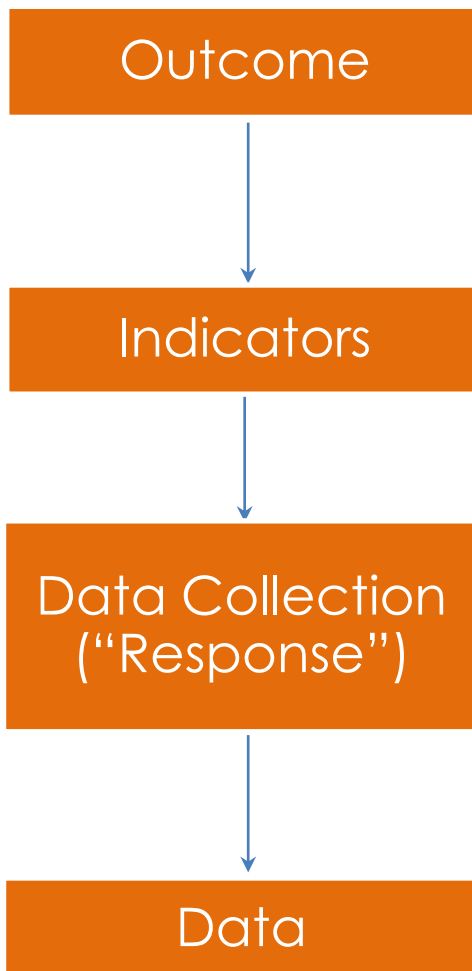


Measuring a concept



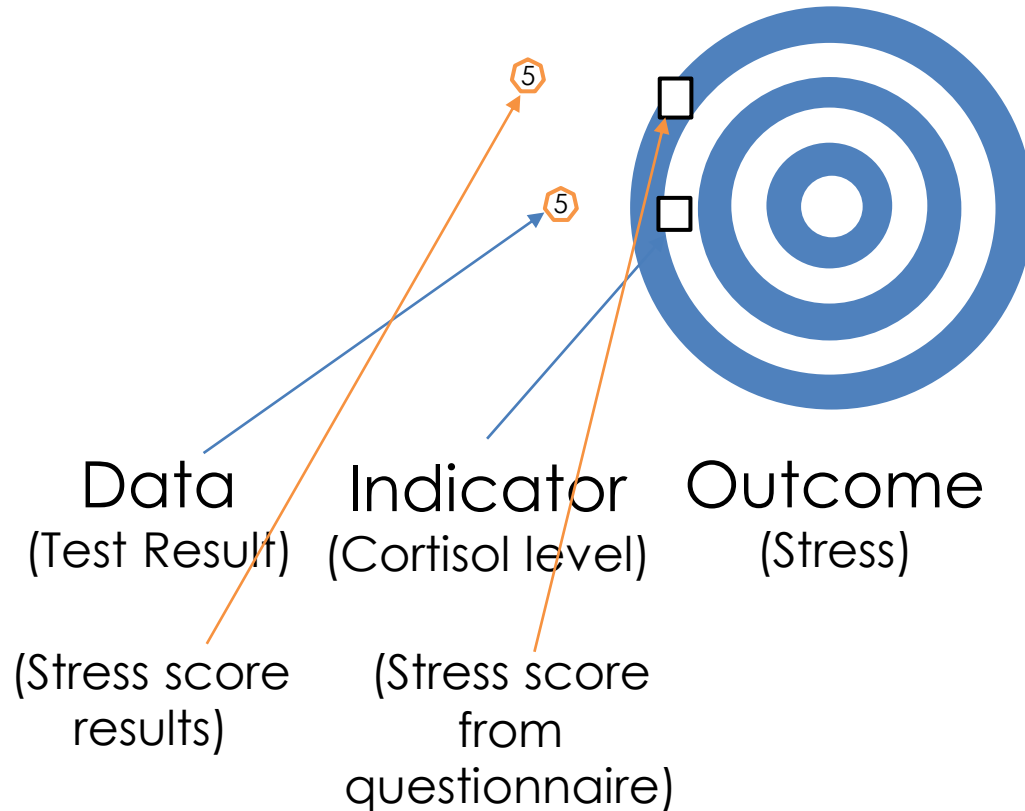
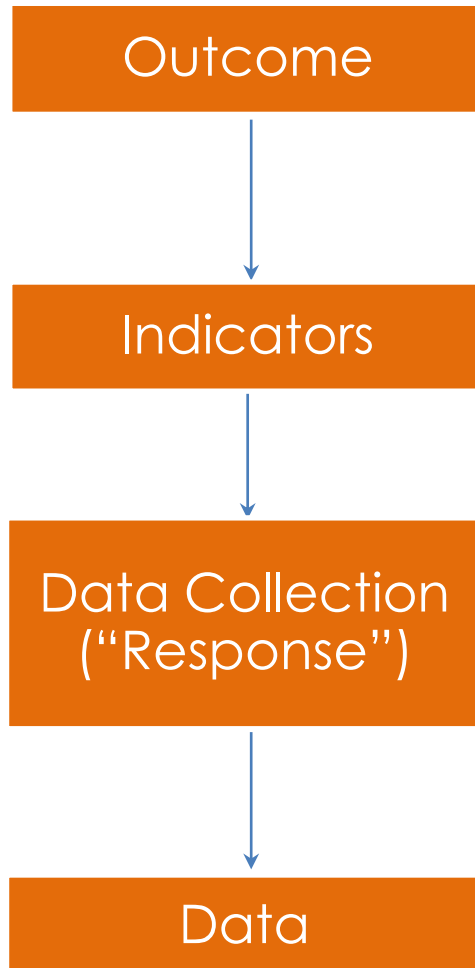
Source: Abdul Jameel Poverty Action Lab

Measuring a concept



Source: Abdul Jameel Poverty Action Lab

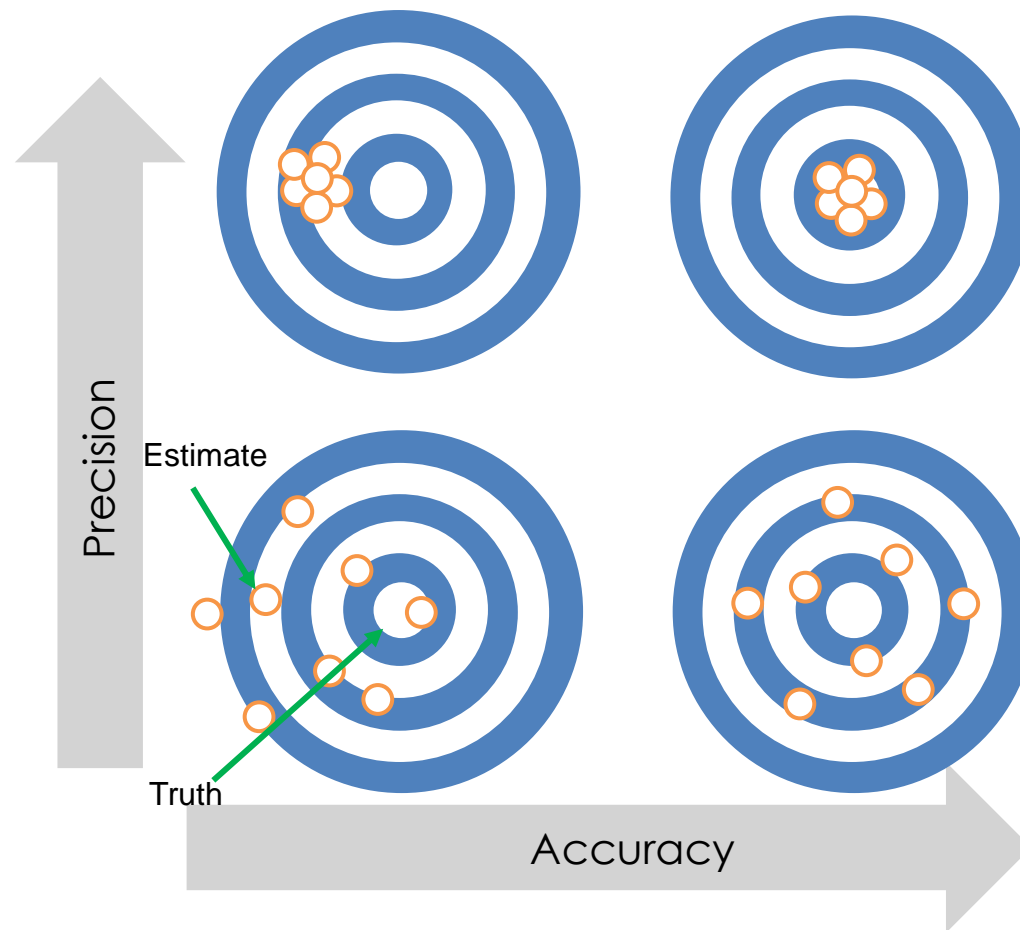
Measuring a concept



Source: Abdul Jameel Poverty Action Lab

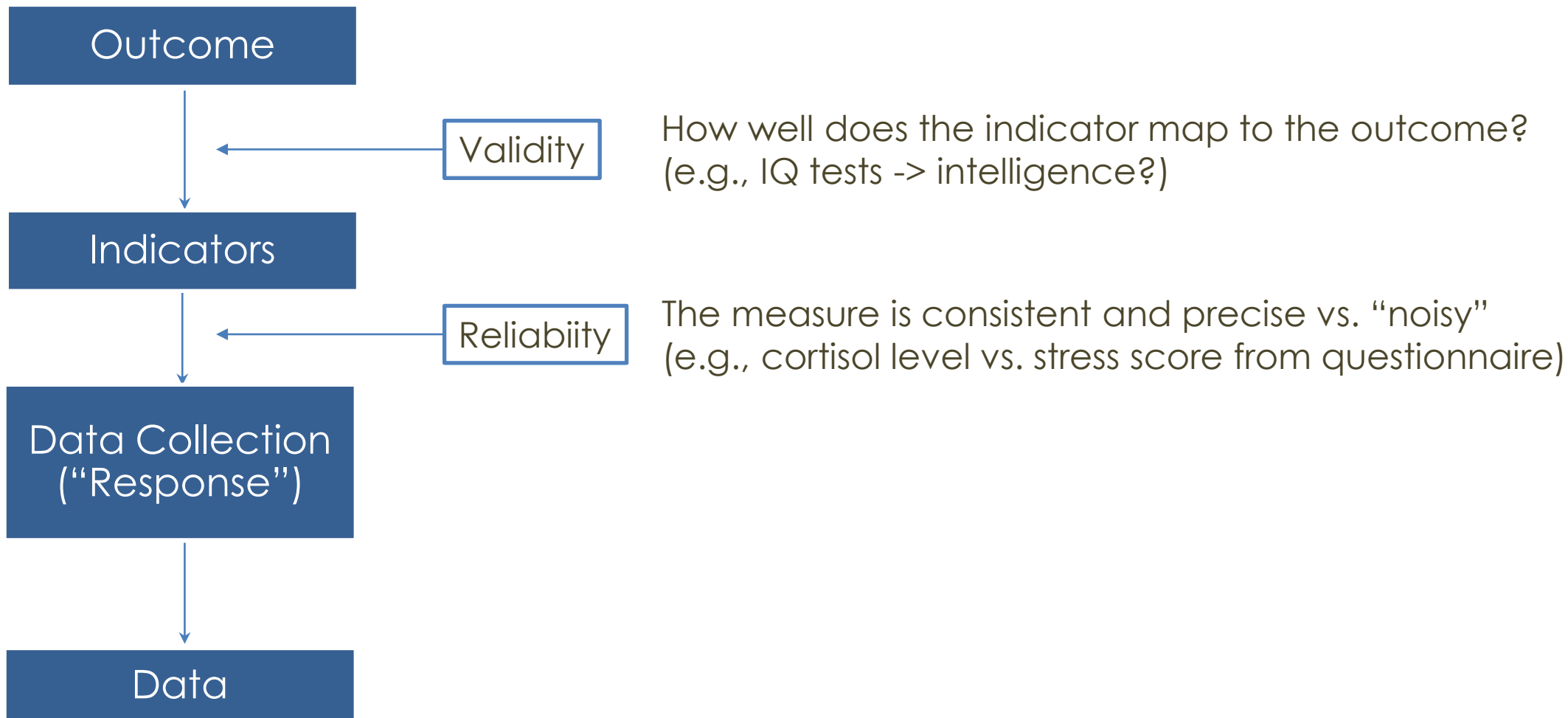
The goals of measurement

- Validity / unbiased
- Accuracy / precision / reliability





The goals of measurement





Data sources – Administrative data

- Evaluate whether the data will be suitable for answering our research questions.
 - Do the data really exist?
 - Have the data been consistently collected?
 - Does the data set cover our population/outcomes of interest?
 - Are the data reliable and unlikely to have been manipulated?
- Limitations:
 - Basic data that cover the target population may not be sufficient.
 - More detailed data are collected only on a random sample of individuals.



Data sources – Own data collection

- In case the required data do not exist or are not reliable or the sample size is not sufficient.
- To create a data set with relevant outcomes for our research questions.
- Requires planning in budget, resources and logistics
- Many trade-offs in deciding what data to collect and when
- Surveys: what to ask, whom, how often (e.g., household expenditure)
- Nonsurvey instruments
 - e.g., tests, direct observation, decision experiments, vignettes, biomarkers, mechanical tracking devices
 - to quantify difficult-to-measure outcomes
 - relatively expensive compared to surveys



Should we conduct a baseline survey?

Worthwhile for

- Limited sample size: use baseline data in the analysis to increase statistical power
- Individual-specific outcomes (e.g., cognitive abilities, test scores, beliefs):
 - highly correlated over time for a given individual
 - With baseline data, a lot of variance between individual can be explained.
- Balance between treatment and control groups:
 - to confirm that randomization was carried out and worked
 - Baseline data are required for stratified randomization.
- subgroup analysis and baseline controls:
 - Subgroup, e.g., by test scores at baseline or income
 - to investigate the effect of a policy on different group of people



Conclusion

- Mapping components of program evaluation (inputs, outputs, outcomes, impacts) guides measurement.
- **Indicators** need to accurately measure.
- Data collection all about *trade-offs*:
 - Quality and cost
 - Validity (accuracy) and reliability (precision)
- Survey / nonsurvey instruments as an option



Thank you!

Questions?