

Implementation science and implementation measurement



Noam Angrist
What Works Hub for
Global Education

What Works Hub for Global Education Conference (WWHGE)

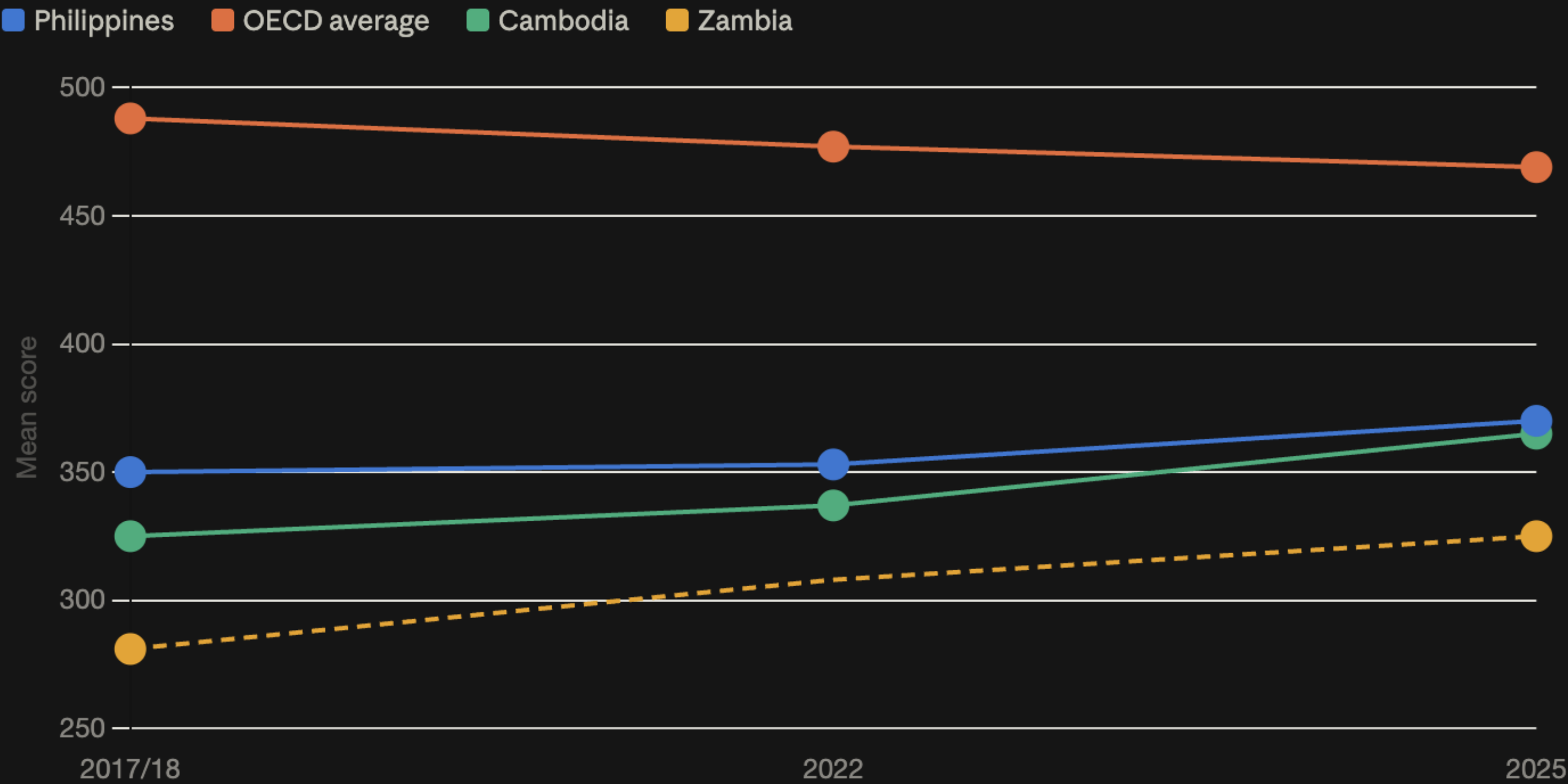
2026

Noam Angrist, Academic Director, What Works Hub for Global Education

Some starting motivation – the learning crisis and some green shoots



PISA Progress (Average Scores)



Learning levels and progress remain alarmingly low and stagnant

Yet, there are some **green shoots** in exactly those countries **prioritizing evidence-based reforms** and with a **growing focus on implementation**

To achieve system-wide gains with limited resources, we need to prioritize “what works” with a laser focus on implementation at scale – and to measure and study these dynamics

Implementation science

Implement the science, and be scientific
about implementation

An implementation science agenda for global education



Science

Current Issue

First release papers

Archive

About ▾

Submit manuscript

HOME > SCIENCE > VOL. 393, NO. 6816 > AN AGENDA FOR IMPLEMENTATION SCIENCE IN EDUCATION

🔒 | POLICY FORUM | EDUCATION



An agenda for implementation science in education

A persistent global learning crisis calls for studying the implementation of “what works” at scale

NOAM ANGRIST, TAHIR ANDRABI, STEFAN DERCON, SHARNIC DJAKER, JAMES HABYARIMANA, MICHELLE KAFFENBERGER, CLARE LEAVER, AKSHAY MANGLA, NOMPUMELELO MOHOHLWANE, AND

ABHIJEET SINGH

fewer

[Authors Info & Affiliations](#)

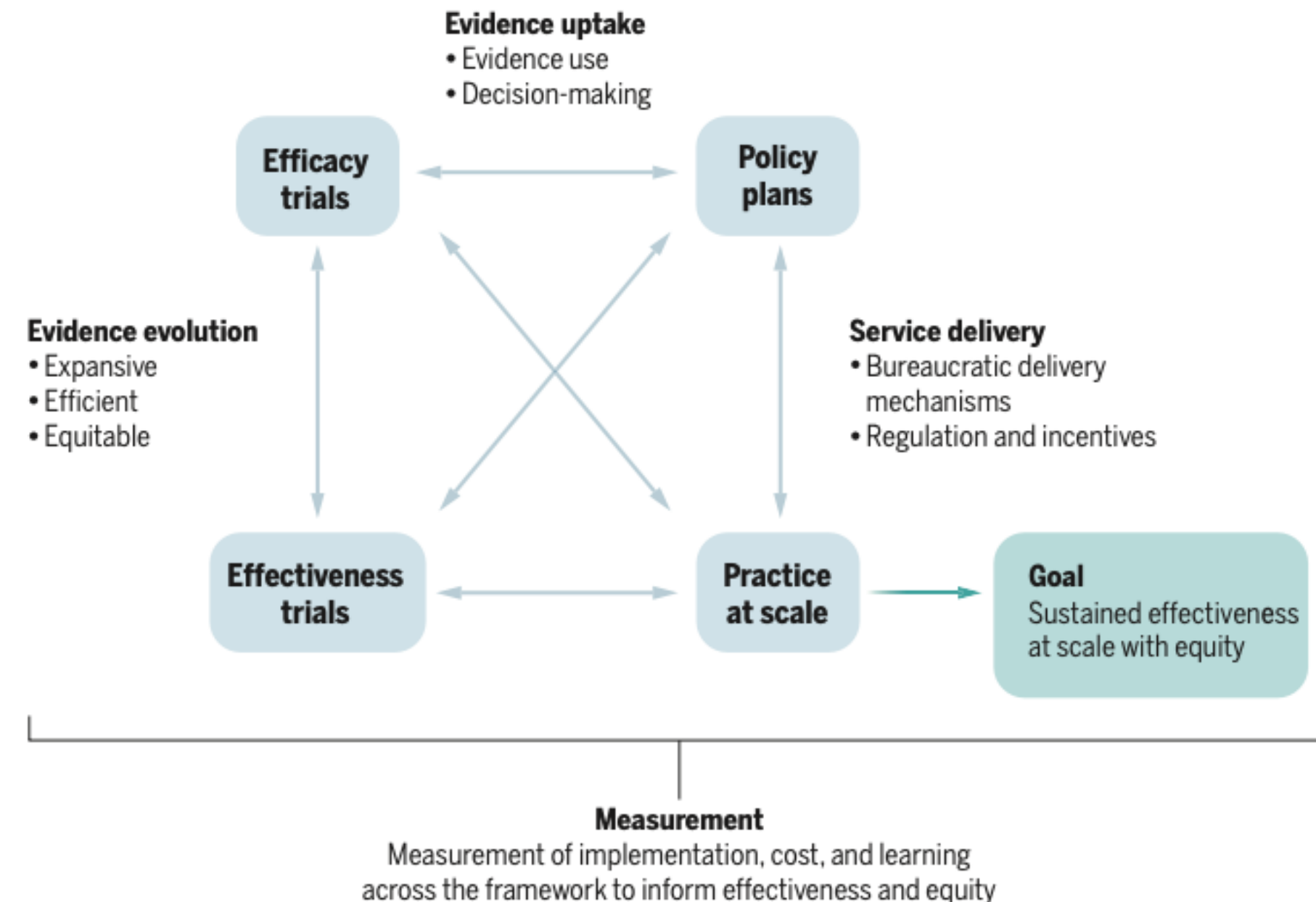
A collective effort of the WWHGE coalition

A simple organizing framework to prioritize areas of advance



A framework for studying implementation in global education

The aim of the framework is to organize, prioritize, motivate, and inform research questions. The framework includes four phases: two in an evidence strand (efficacy trials to effectiveness trials), two in a policy and practice strand (policy plans to practice at scale), and connections between them. Policies and programs might move through the framework in different ways, or require multiple iteration cycles.



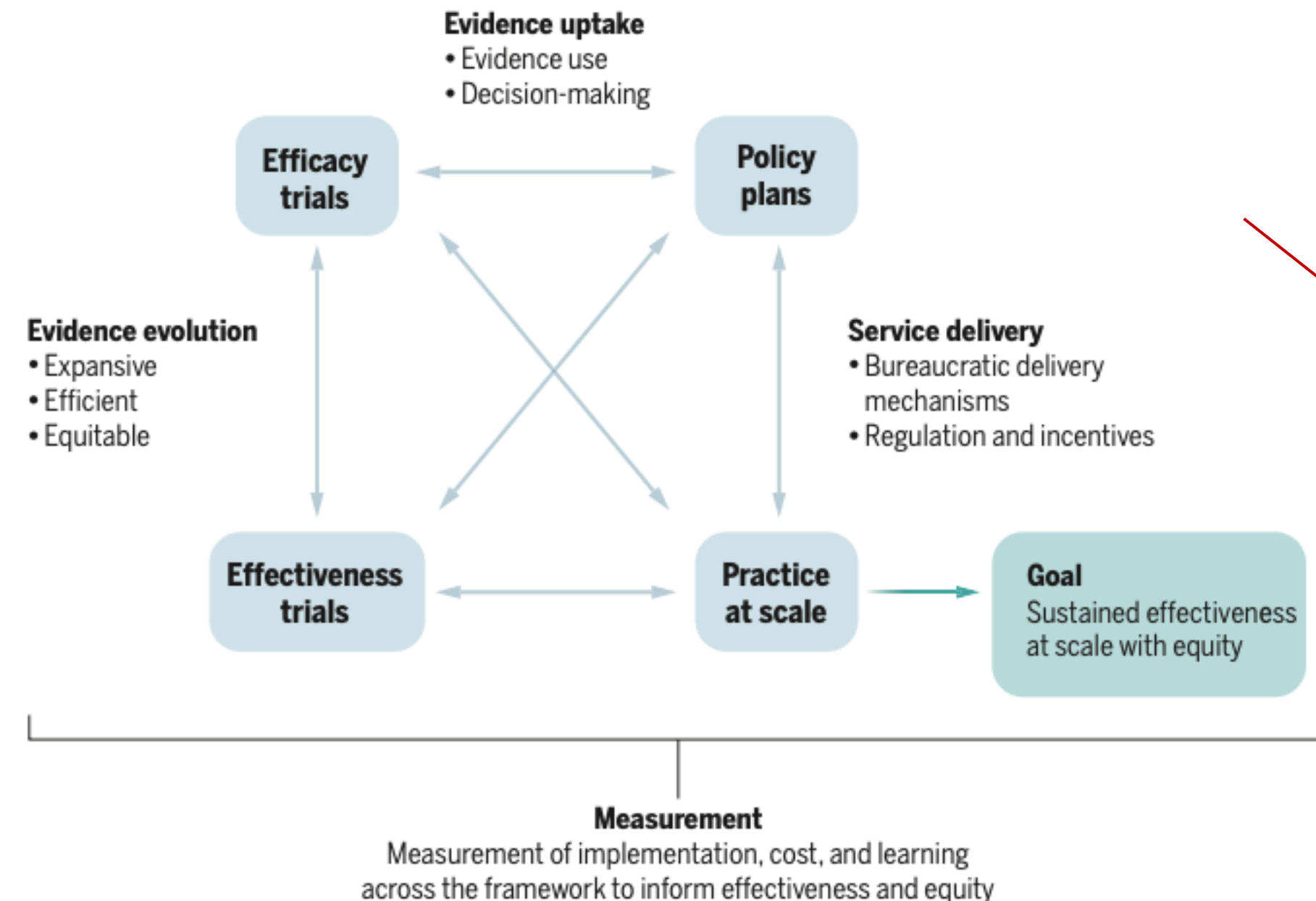
The bubbles are back! And better than ever...

An animating goal is to prevent or reverse “voltage drops”



A framework for studying implementation in global education

The aim of the framework is to organize, prioritize, motivate, and inform research questions. The framework includes four phases: two in an evidence strand (efficacy trials to effectiveness trials), two in a policy and practice strand (policy plans to practice at scale), and connections between them. Policies and programs might move through the framework in different ways, or require multiple iteration cycles.



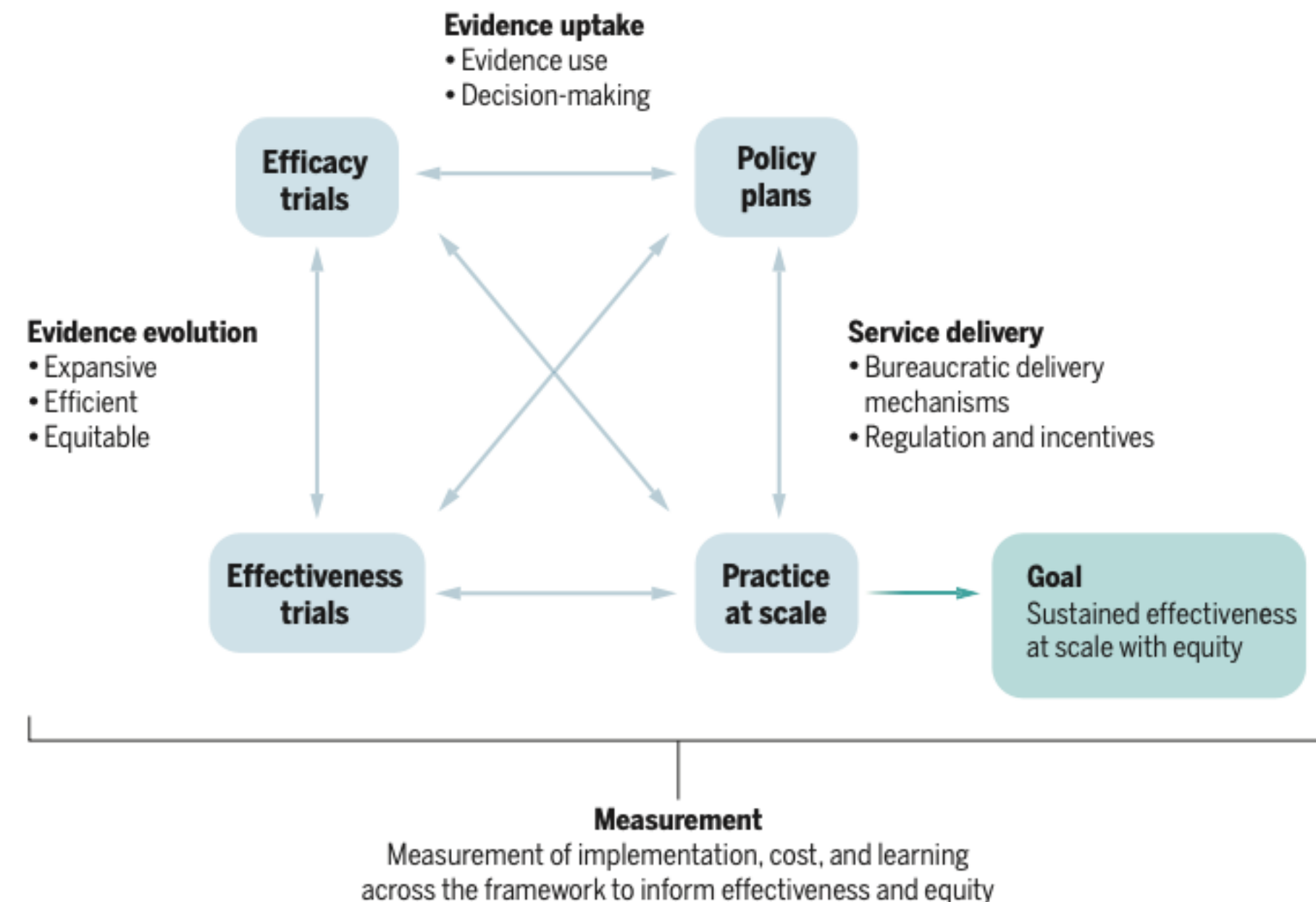
Voltage drops

Focused on the leap from efficacy → effectiveness trials



A framework for studying implementation in global education

The aim of the framework is to organize, prioritize, motivate, and inform research questions. The framework includes four phases: two in an evidence strand (efficacy trials to effectiveness trials), two in a policy and practice strand (policy plans to practice at scale), and connections between them. Policies and programs might move through the framework in different ways, or require multiple iteration cycles.



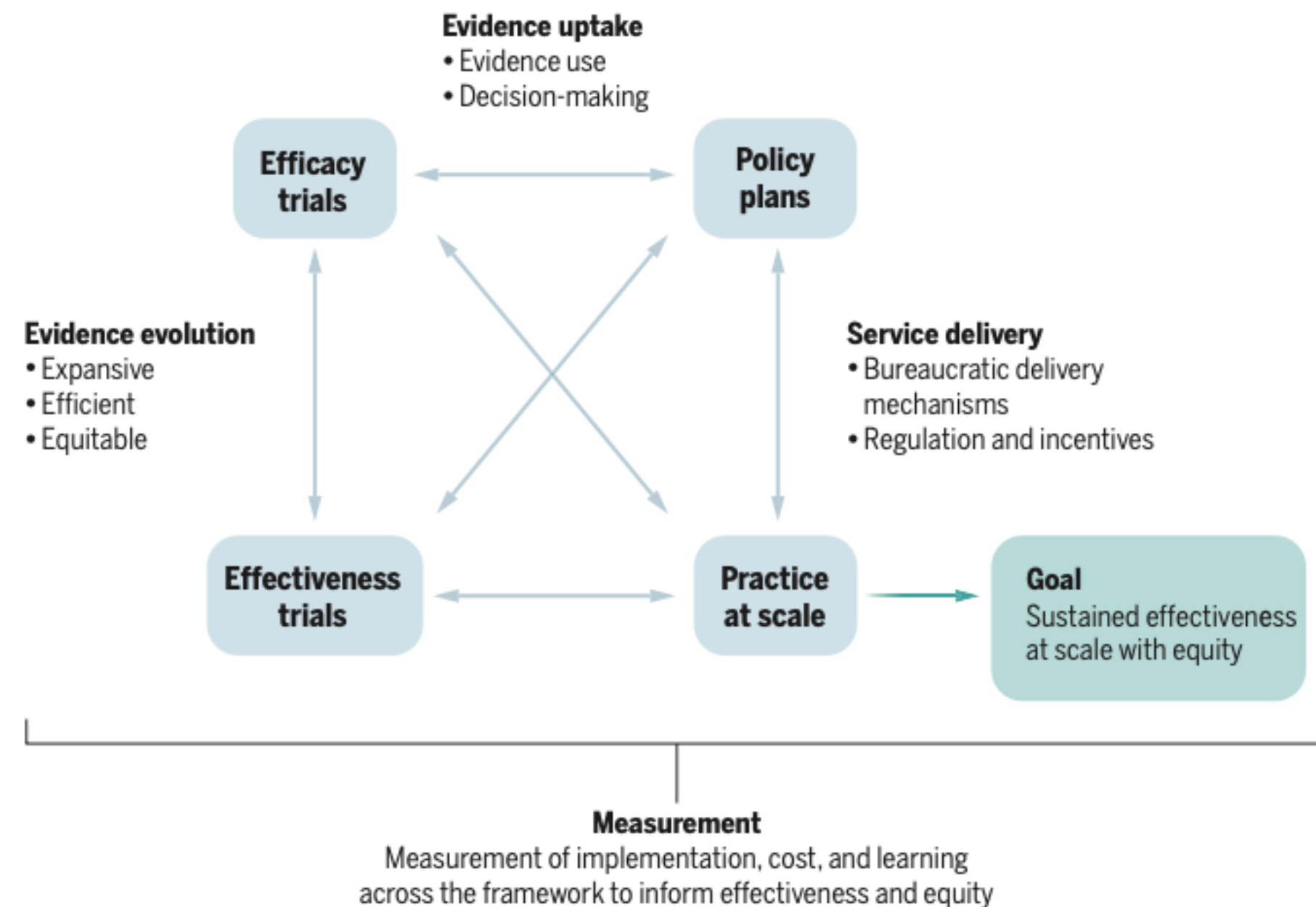
Some concrete approaches:

- Testing scalable models
- Adaptive testing
- Multi-context studies

Focused on the leap from efficacy → policy & practice



What type of evidence is most persuasive to decisionmakers?

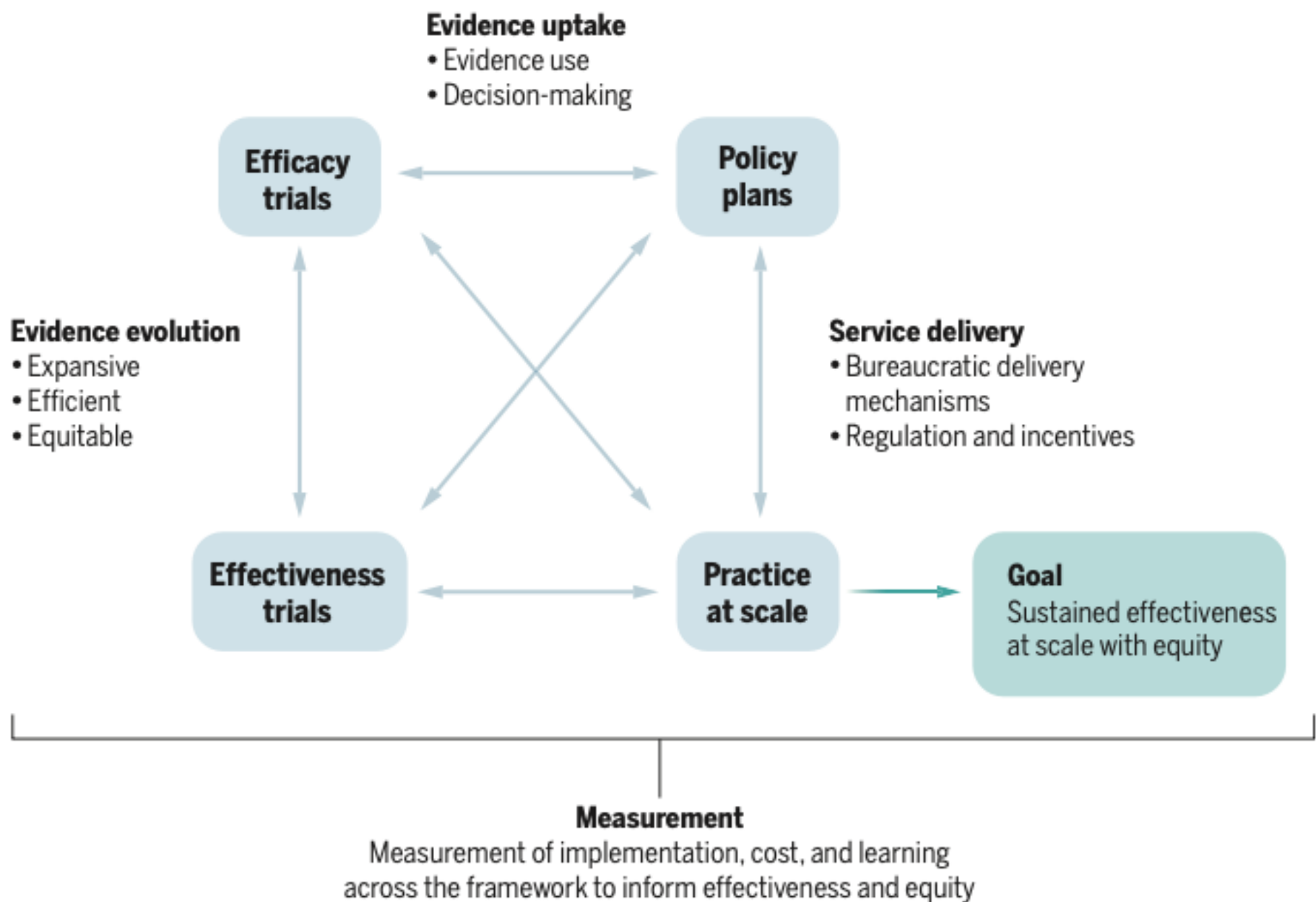


Focused on the leap from policy → practice



A framework for studying implementation in global education

The aim of the framework is to organize, prioritize, motivate, and inform research questions. The framework includes four phases: two in an evidence strand (efficacy trials to effectiveness trials), two in a policy and practice strand (policy plans to practice at scale), and connections between them. Policies and programs might move through the framework in different ways, or require multiple iteration cycles.



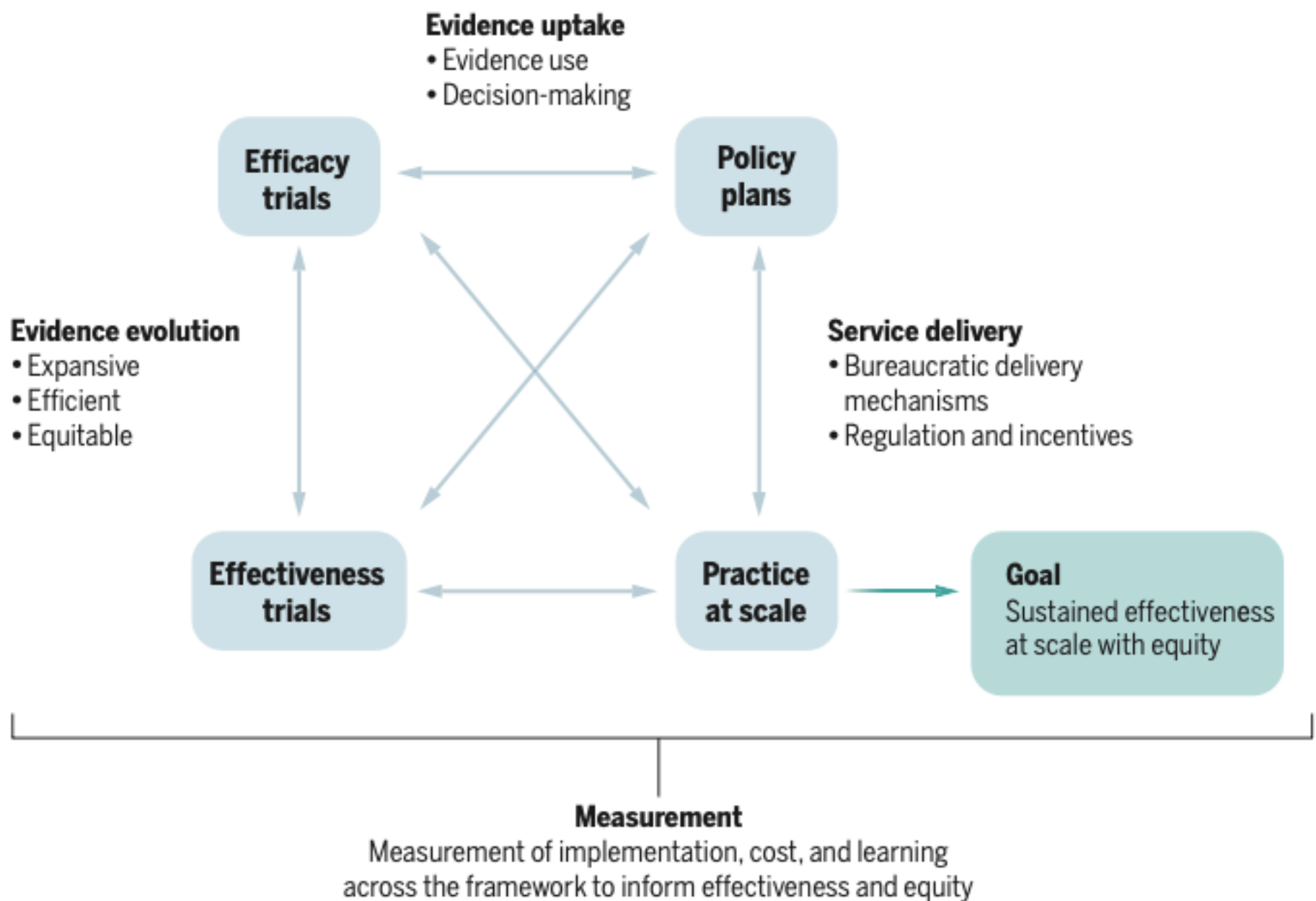
These gaps are larger than often assumed – we need to find ways to close them

Focused on the leap from policy → practice



A framework for studying implementation in global education

The aim of the framework is to organize, prioritize, motivate, and inform research questions. The framework includes four phases: two in an evidence strand (efficacy trials to effectiveness trials), two in a policy and practice strand (policy plans to practice at scale), and connections between them. Policies and programs might move through the framework in different ways, or require multiple iteration cycles.



Aim for credible & low-cost Shared reporting standards

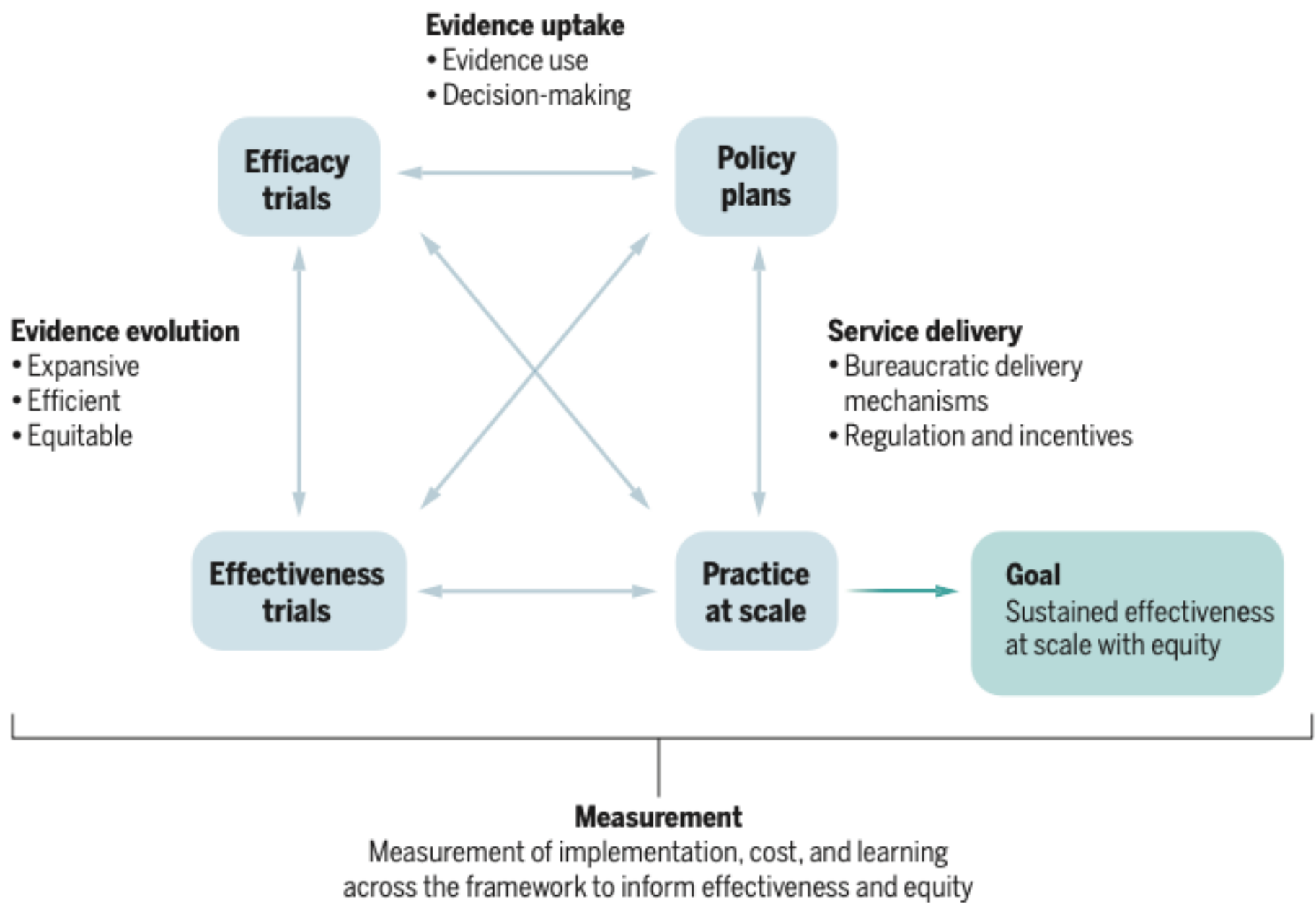
A living agenda for continued
reflection and refinement

Zooming in on measurement



A framework for studying implementation in global education

The aim of the framework is to organize, prioritize, motivate, and inform research questions. The framework includes four phases: two in an evidence strand (efficacy trials to effectiveness trials), two in a policy and practice strand (policy plans to practice at scale), and connections between them. Policies and programs might move through the framework in different ways, or require multiple iteration cycles.



Aim for credible & low-cost Shared reporting standards

As if one framework wasn't enough...
now another framework on
measuring implementation

First, some motivation

Less than 13% of impact evaluations currently quantify and formally account for implementation.

Measurement is the foundation for understanding and progress.

Measurement Concepts

Mapping the education system with stages, units, and agents

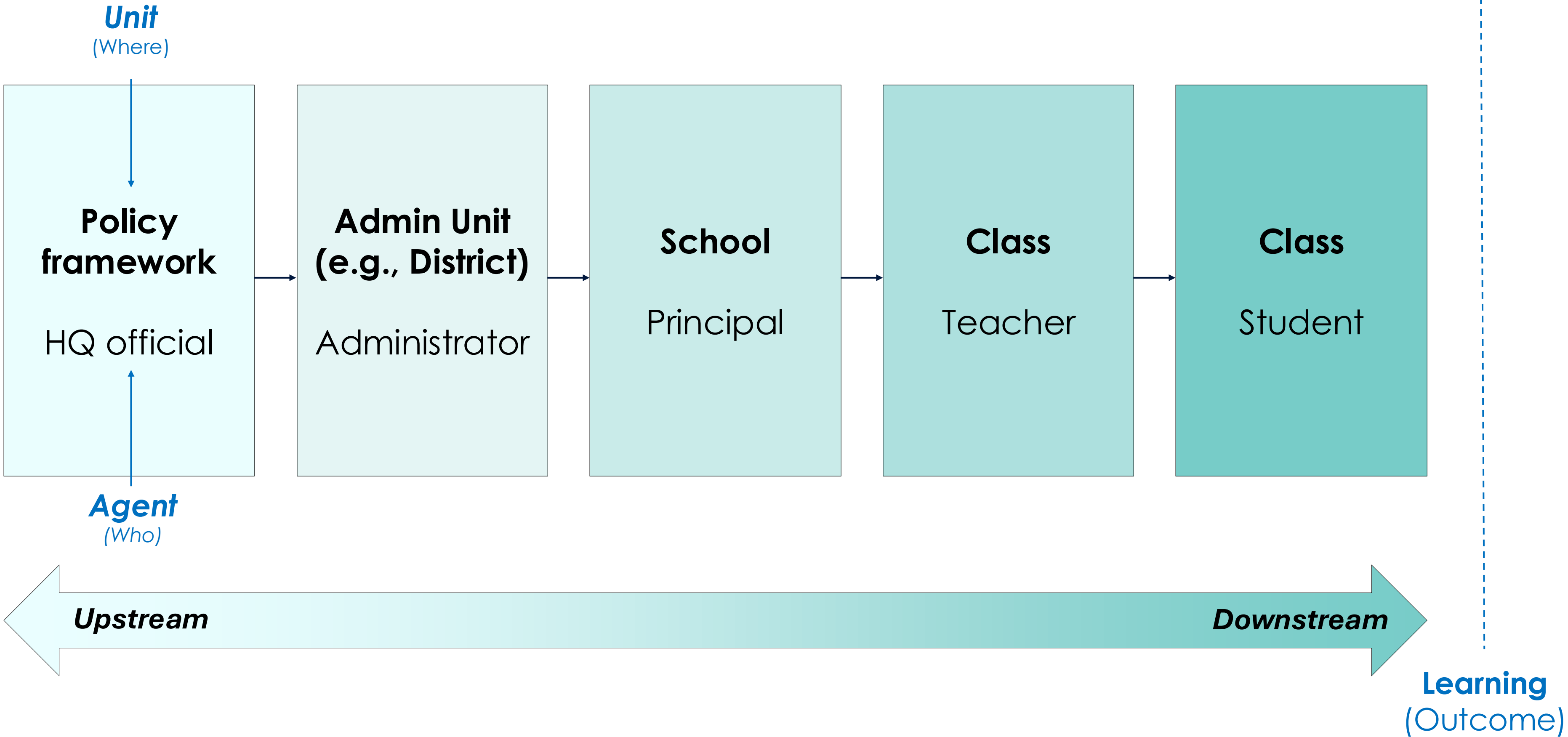
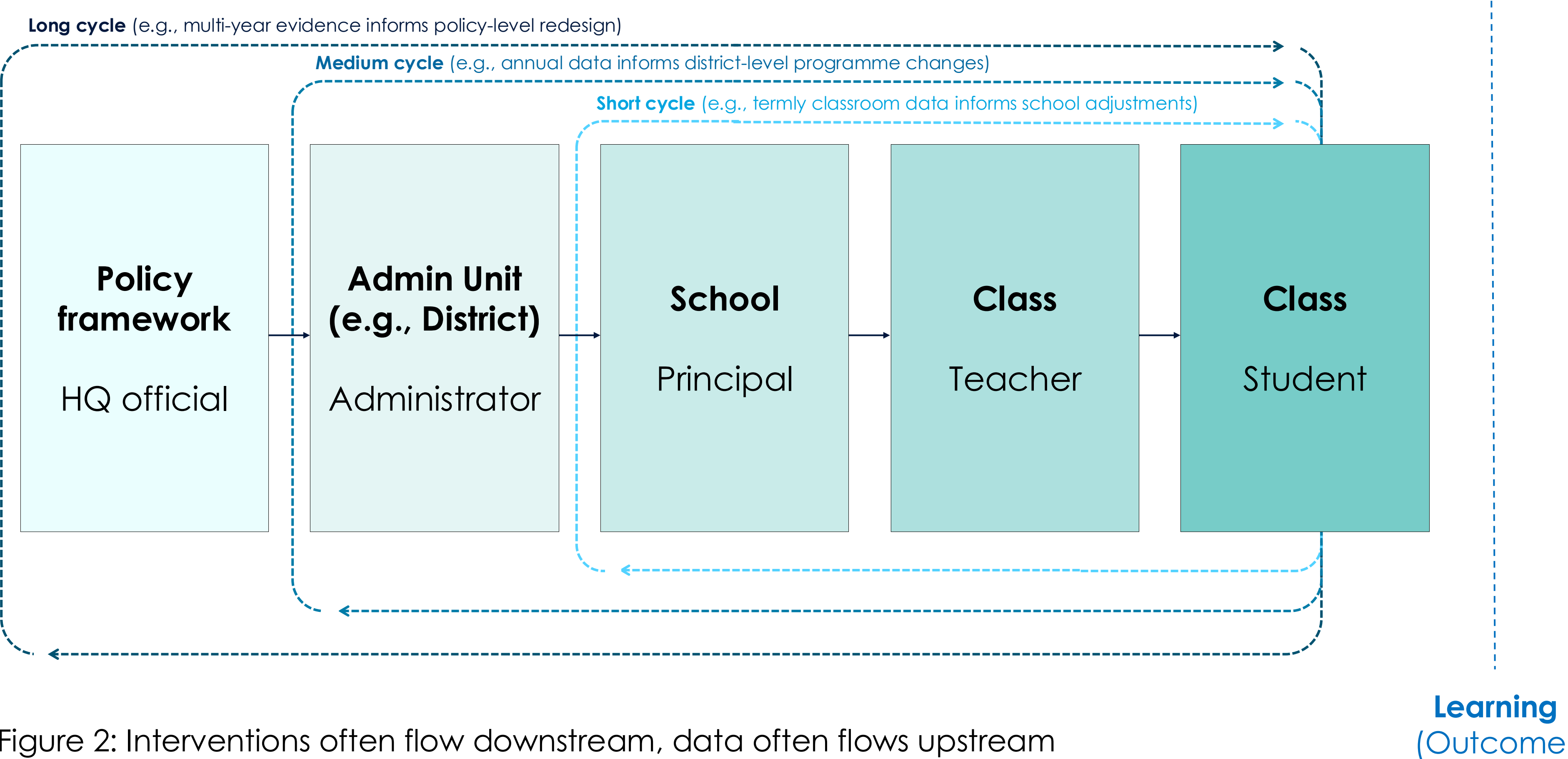
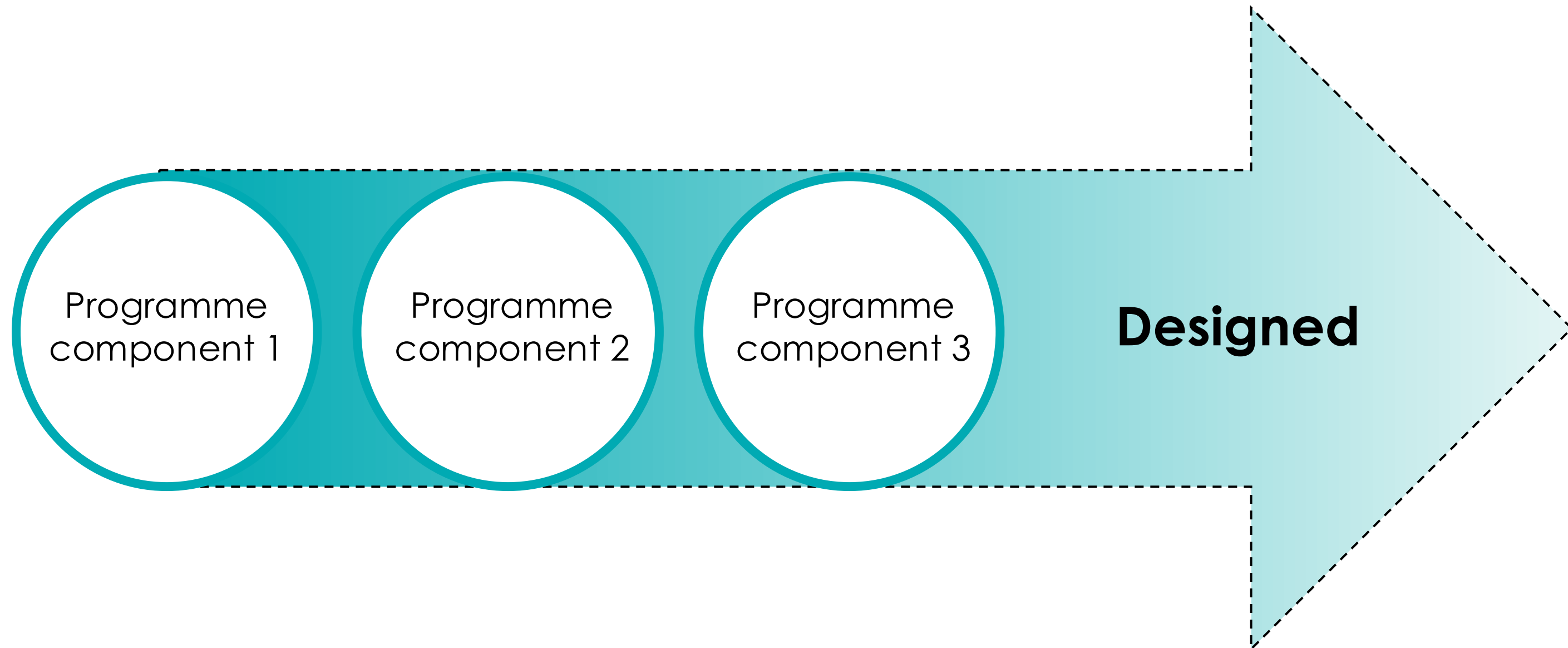


Figure 1: Implementation stages, agents and units in an education system

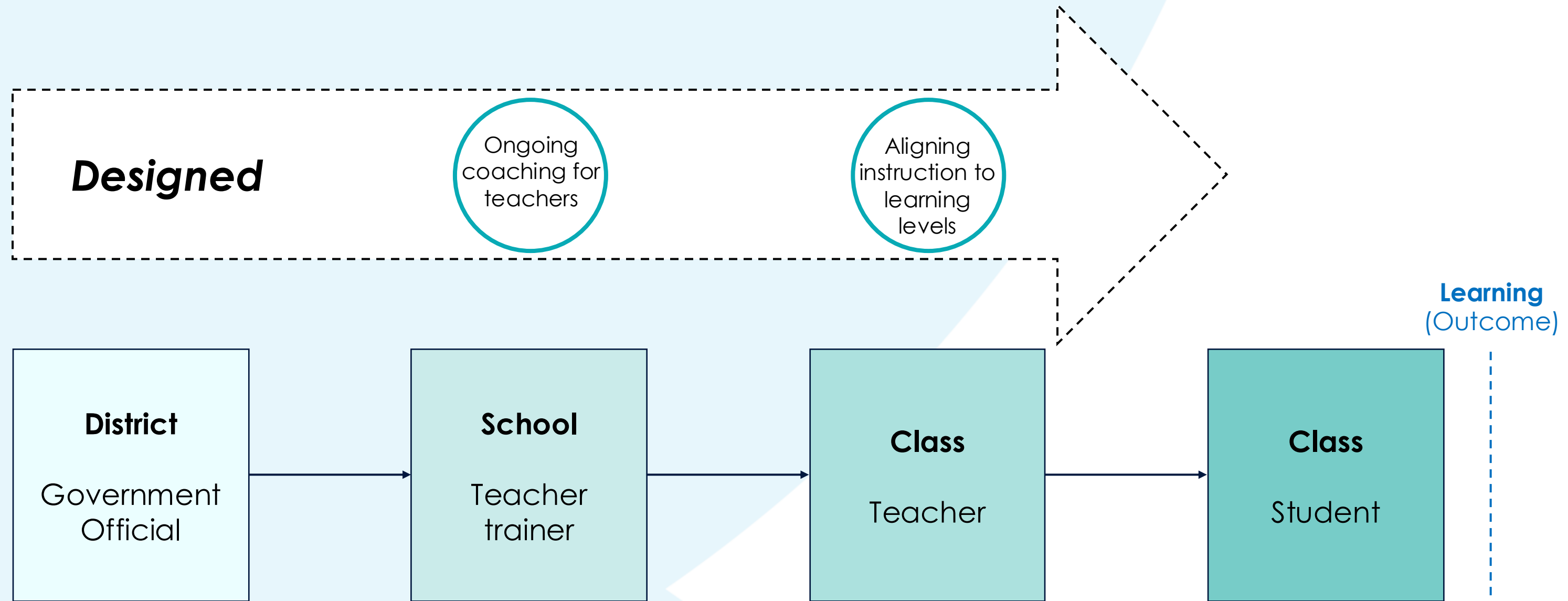
Interventions often flow down, data flows up in short vs long cycles



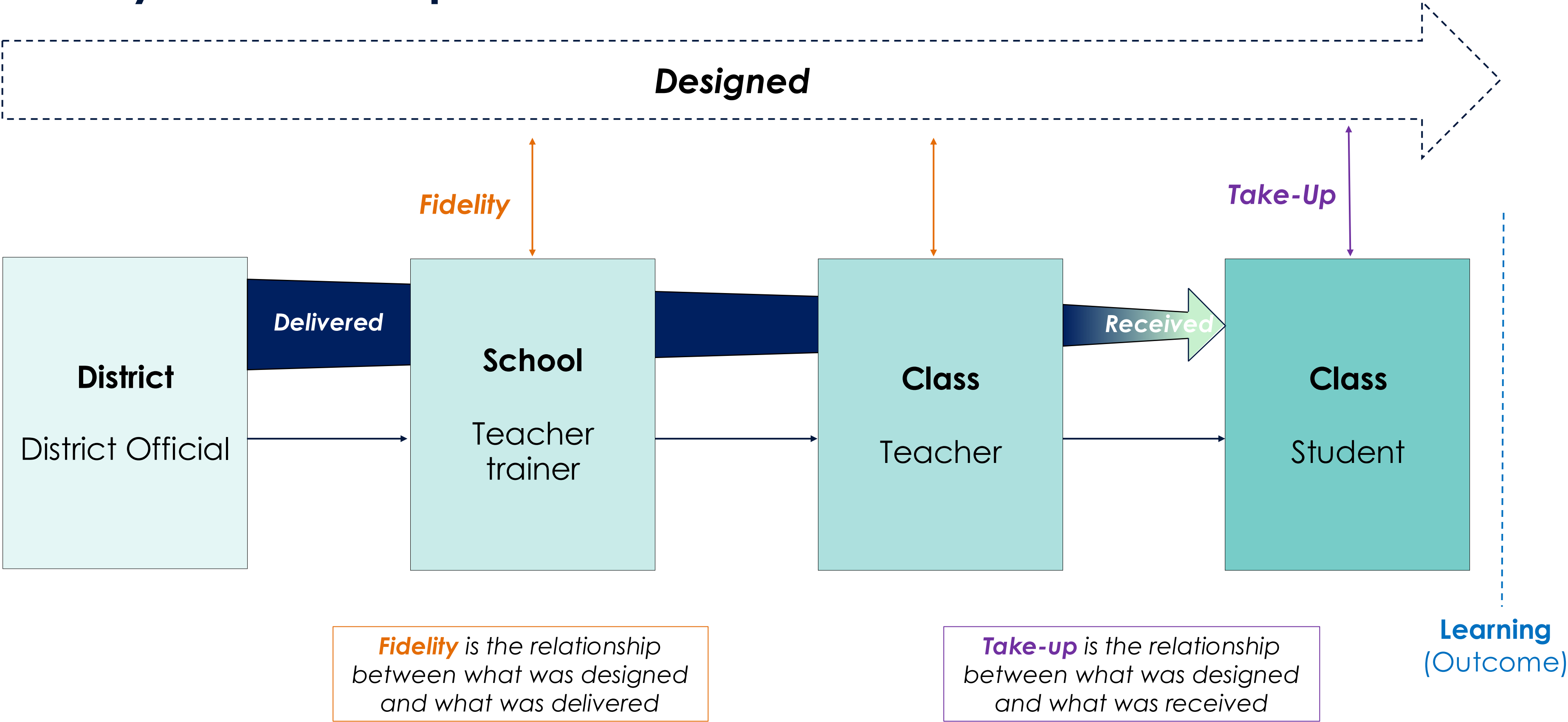
Components are prioritized in the design phase



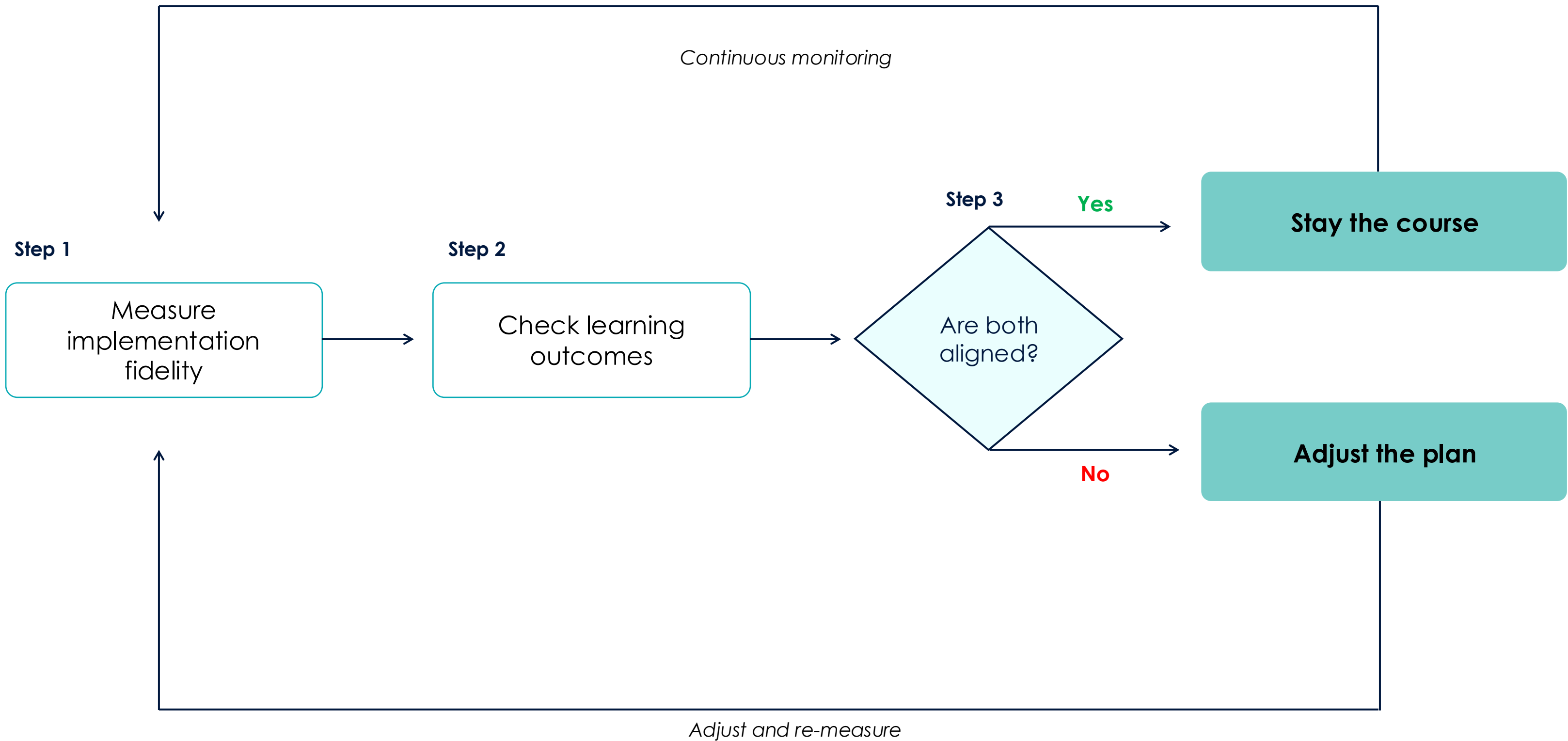
Components are situated in the system



Fidelity and take-up

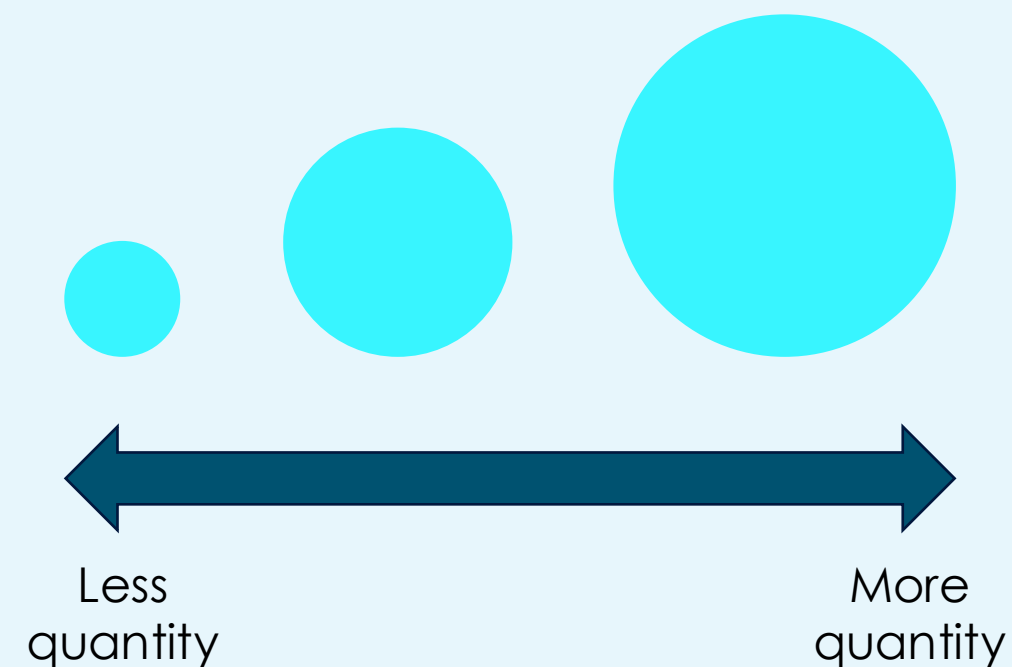


Implementation fidelity is only “good” if it leads to good outcomes

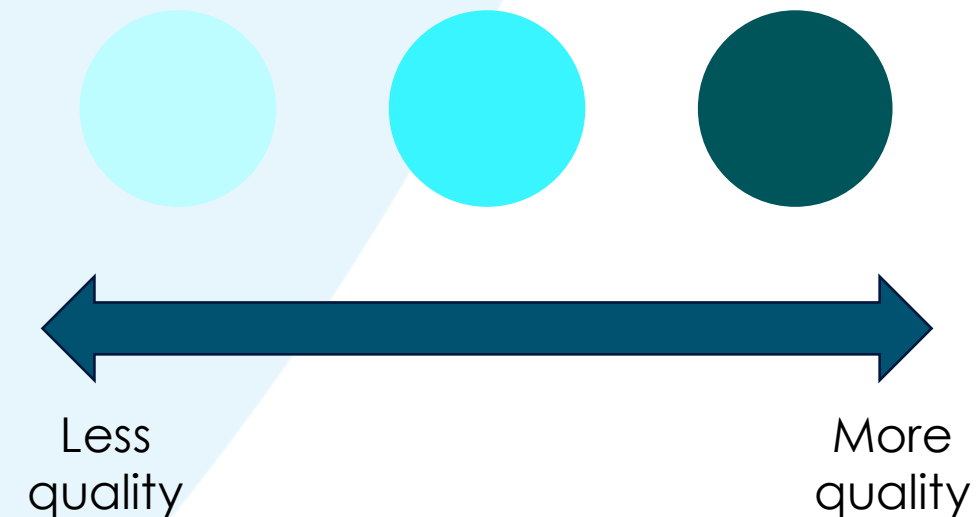


Measurement Dimensions

Two key dimensions of implementation fidelity: quantity and quality



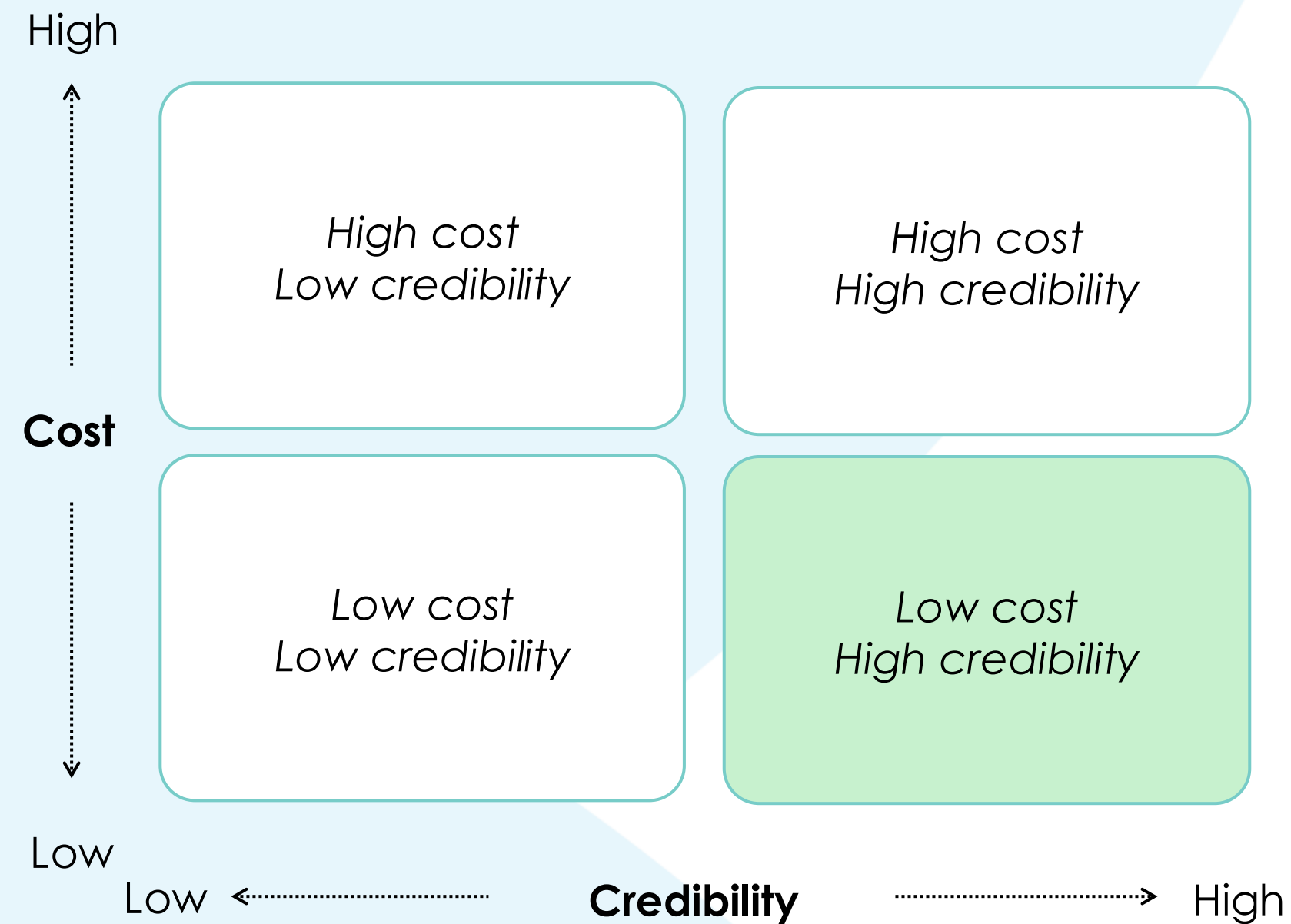
Measures of **quantity** capture how *much* of something was delivered (e.g., how many hours of instruction, how many textbooks)



Measures of **quality** capture how *well* something was delivered (e.g., were children grouped by level correctly, did the teacher stick to the lesson plan)

Operationalizing measurement
through modes and metrics

The measurement “sweet spot”: rigor meets practicality



Credibility Analysis



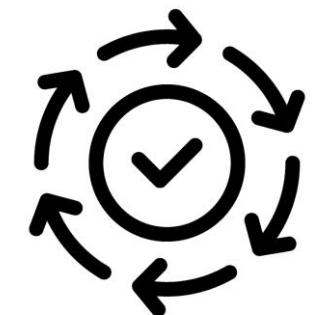
Vital Signs – health checks on the data, is there a pulse?

- Ceiling effects
- Flooring effects
- Variance
- Missingness
- Sample size



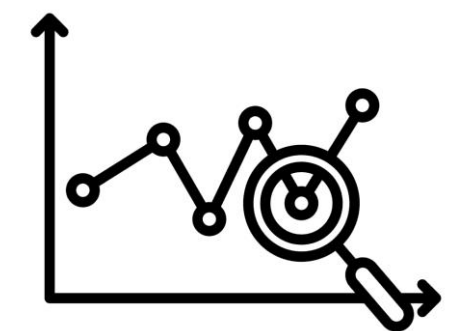
Construct validity

- Are related indicators correlated and telling a similar story



Predictive validity

- Are indicators linked to downstream outcomes of interest (e.g. learning)



Measurement modes



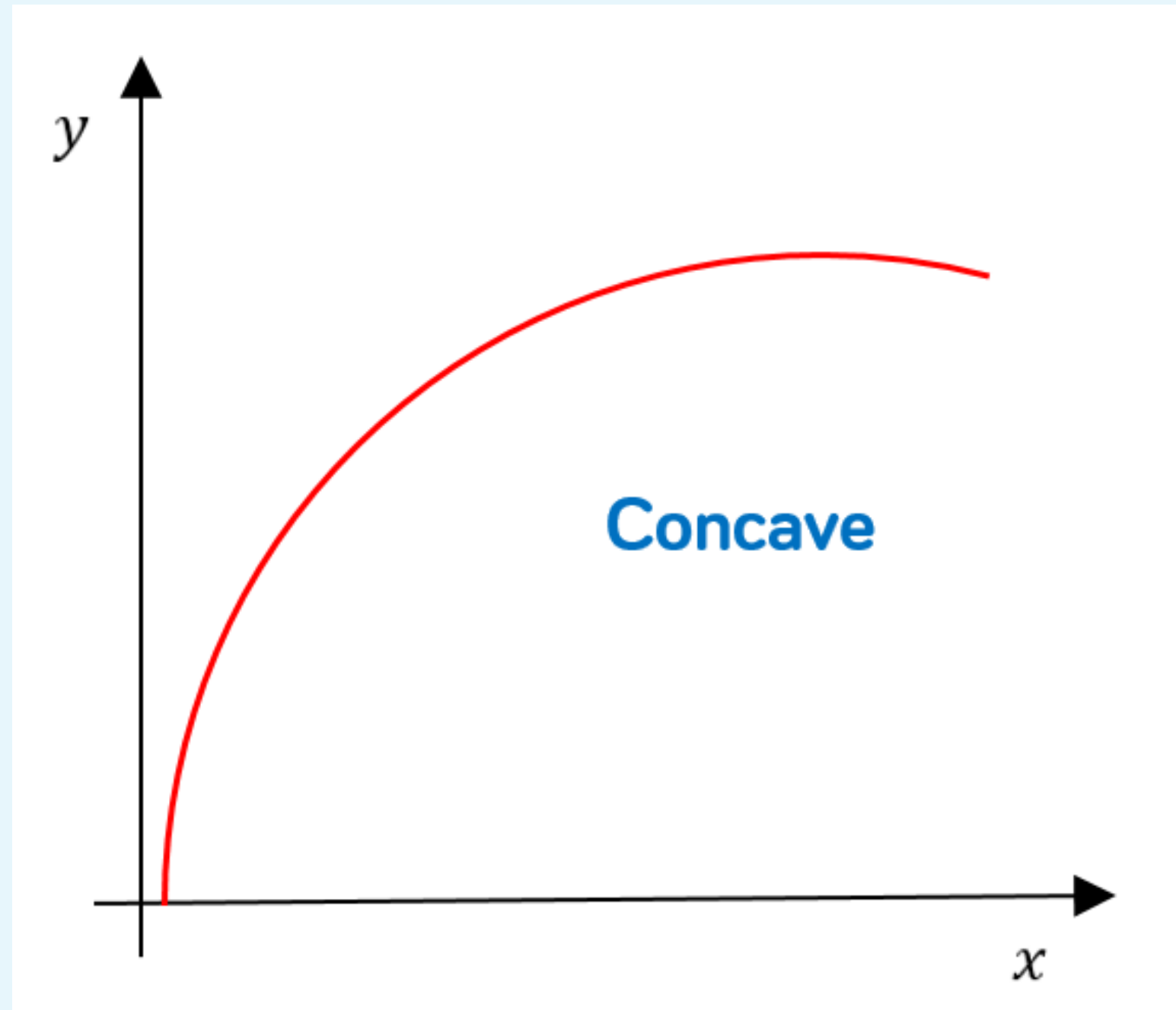
Tool Type	Cost	Credibility	Good for
In-person survey	Medium	High	In-depth, credible information
Phone survey	Low	Medium	Monitoring implementation for real-time improvement
Administrative data	Low	Medium	Monitoring implementation for real-time improvement, when relevant data are collected and reliable data systems exist
External observation (e.g., class observations)	High	High	In-depth, credible information; verifying other modes of data collection

Measurement Use Cases

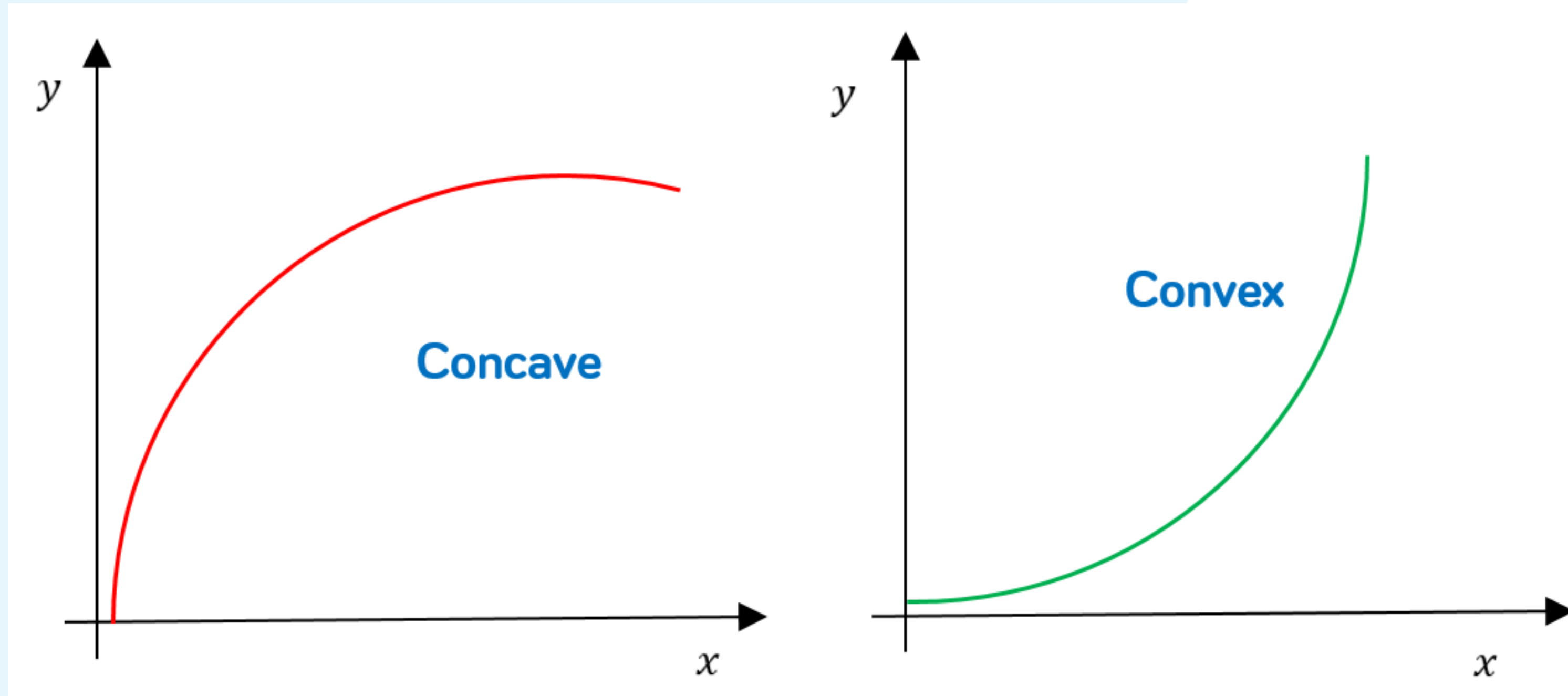
Use case 1:

Real-time adaptive testing using
downstream quality implementation
indicators

Can implementation measurement help reverse the voltage drop?



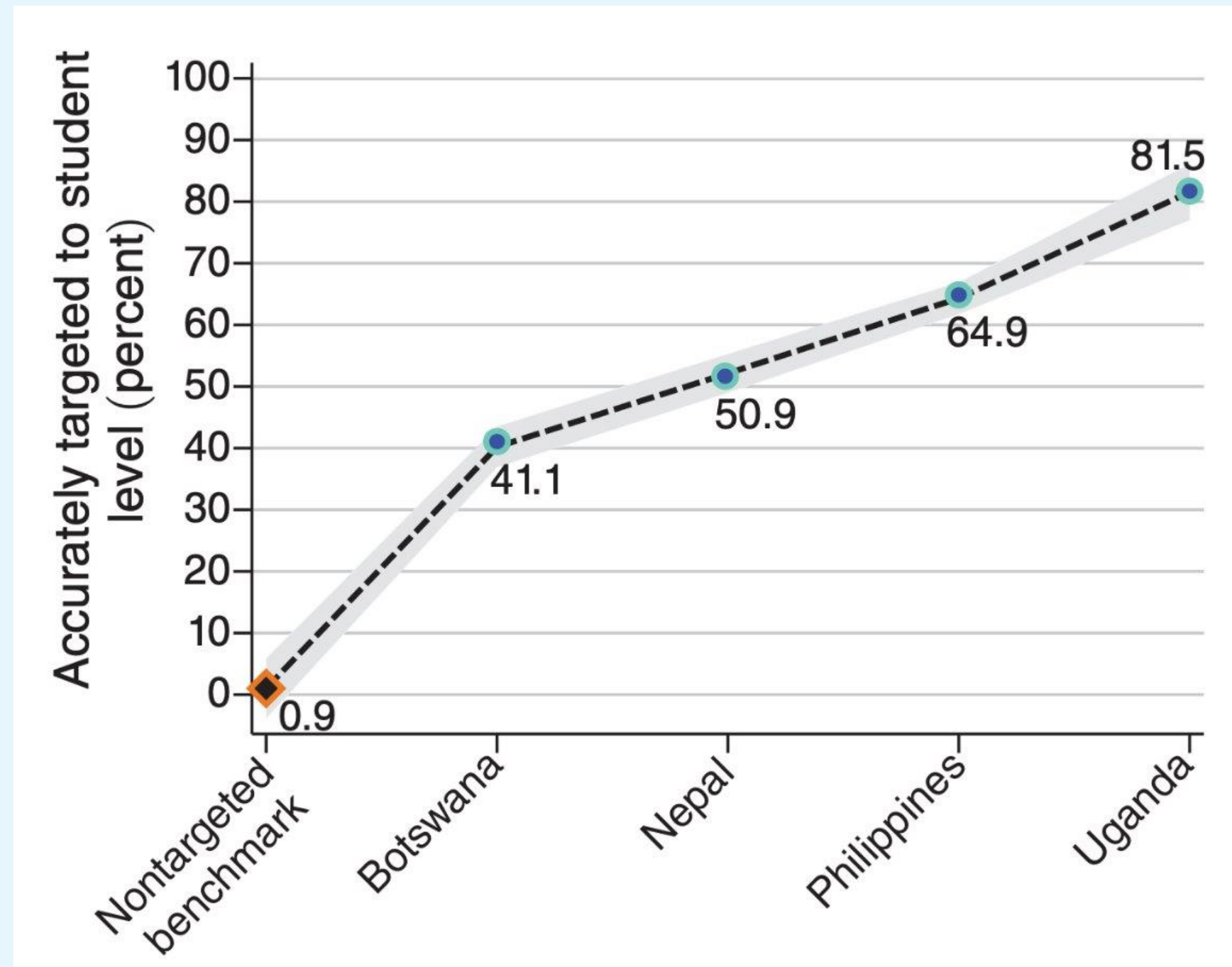
Can implementation measurement help reverse the voltage drop?



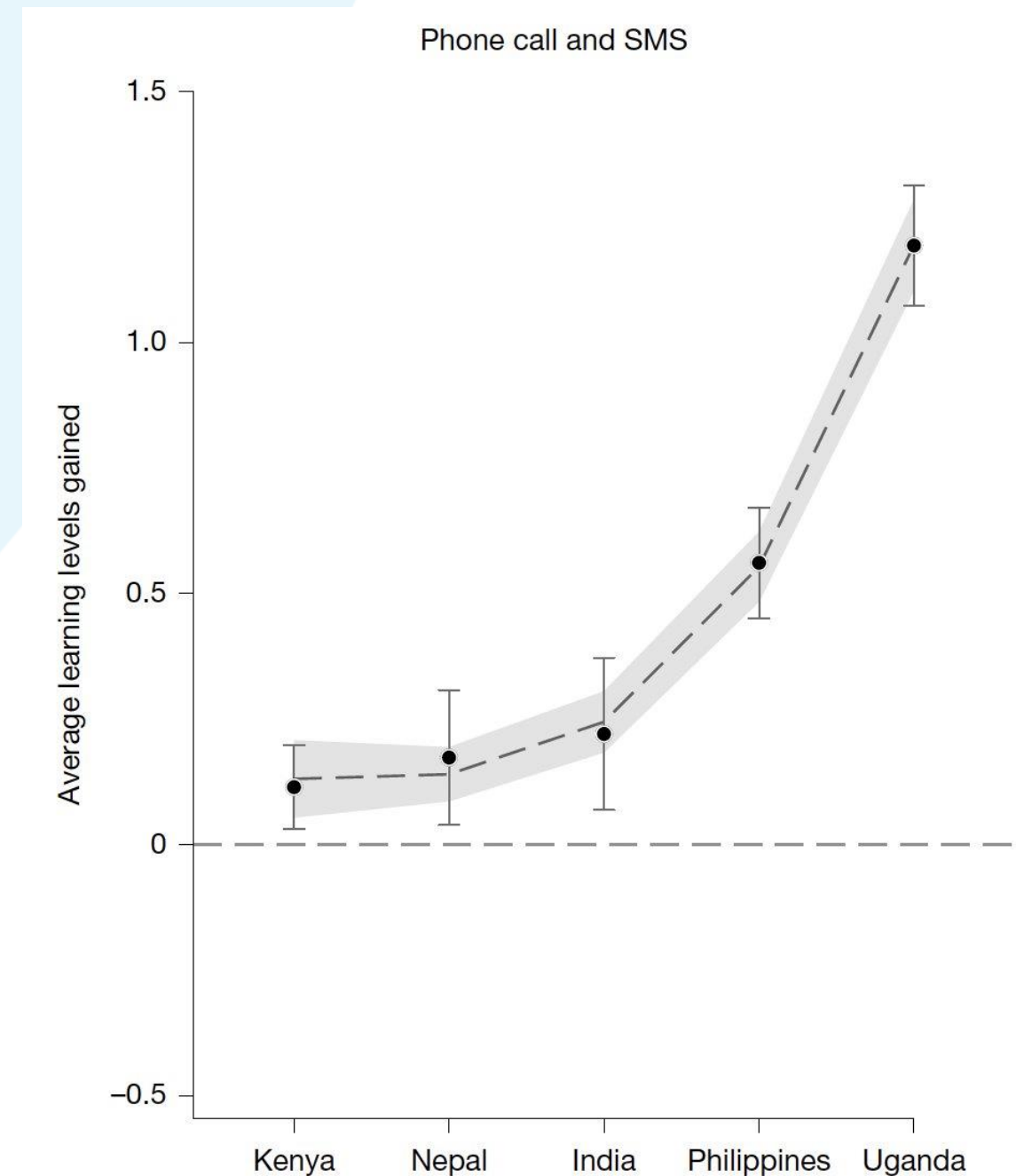
A use case to course-correct real-time



Implementation fidelity improves over time



Learning improves over time

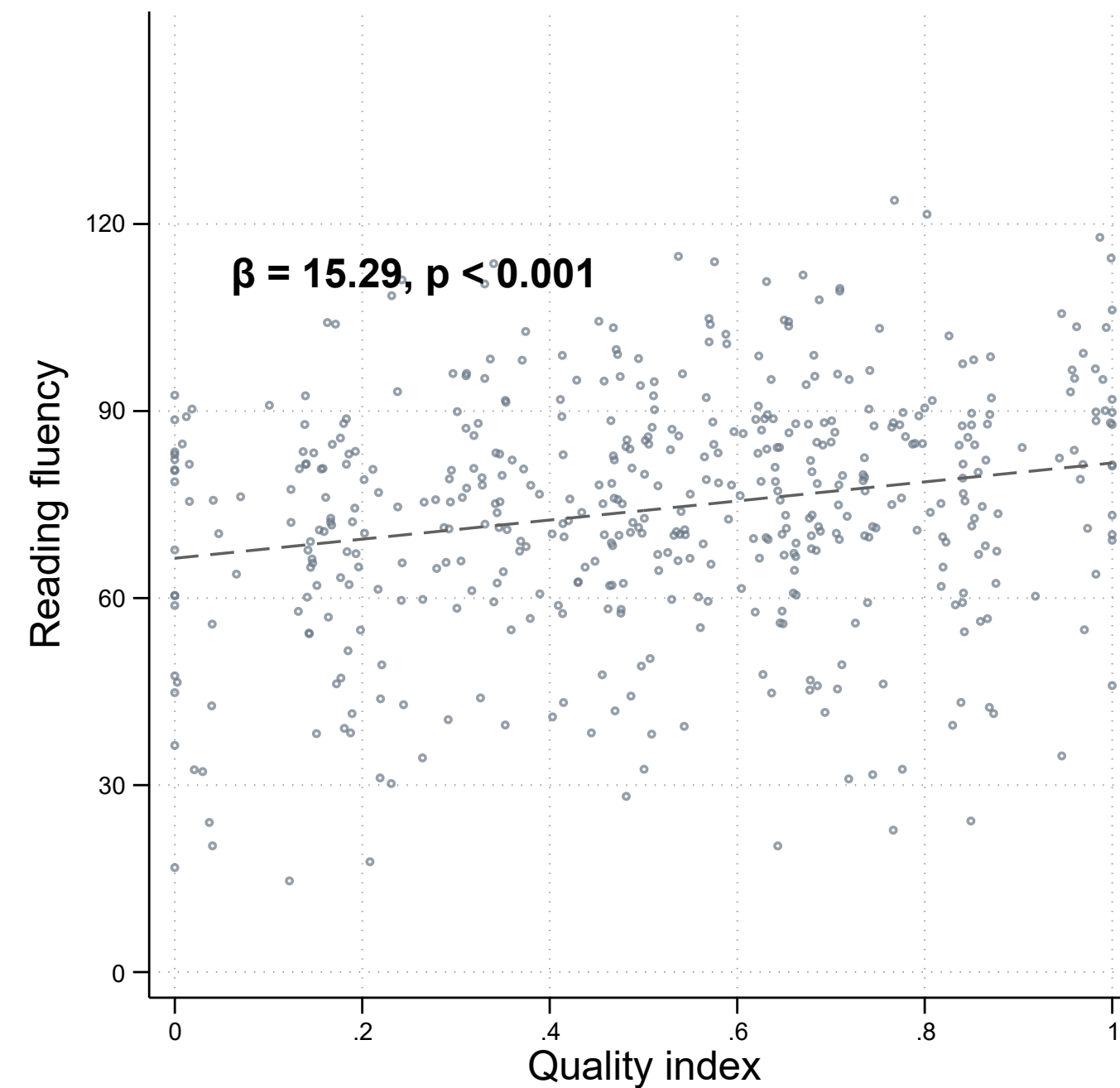
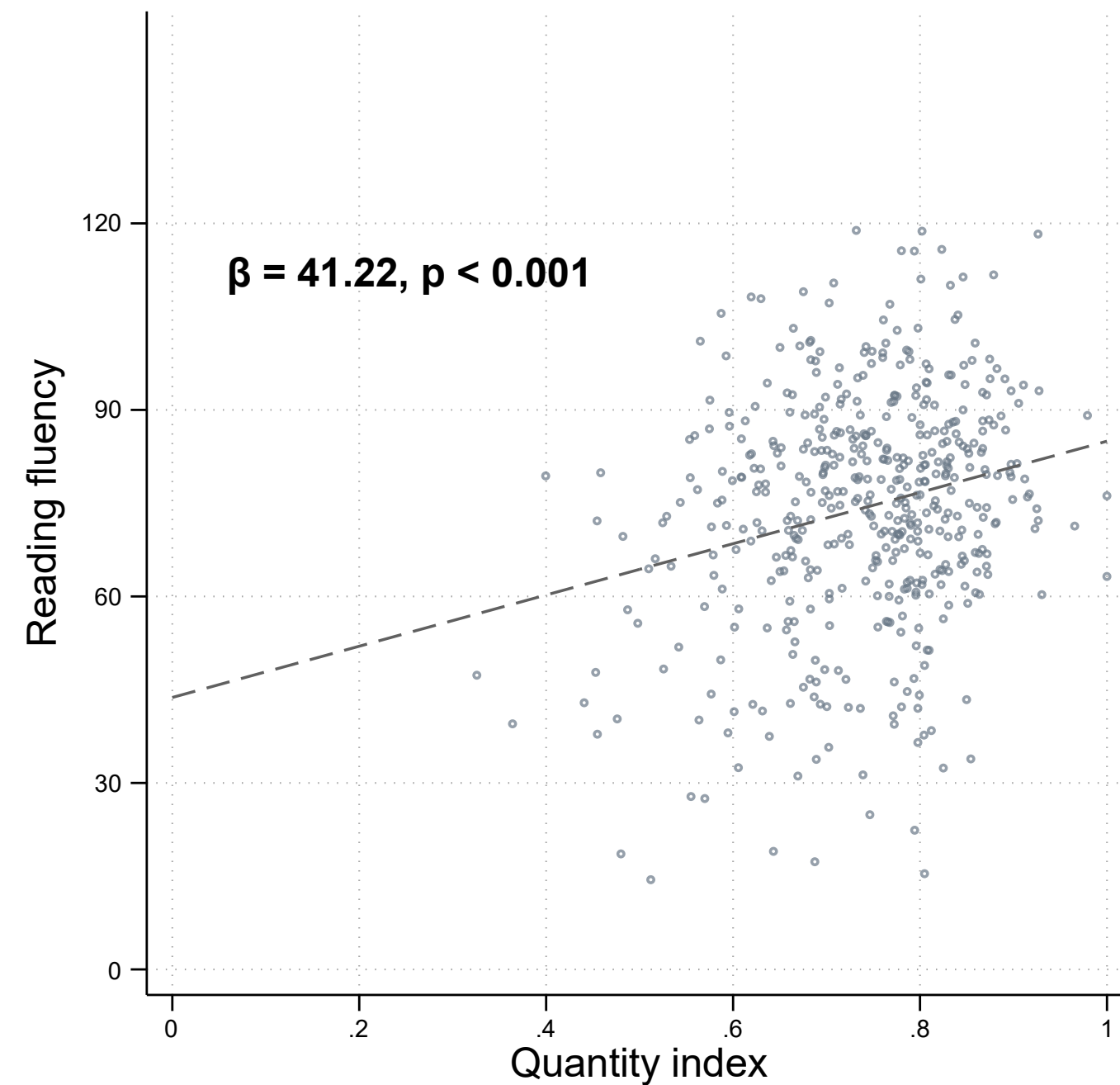


Implementation fidelity and learning levels improve by country and over time (adapted from Angrist et al., 2026) in *Nature*

Use case 2:

Comparing quantity vs. quality to identify the measurement sweet spot of low-cost, high credibility indicators

Quantity shows stronger links with learning, and is low cost – more in the next panel



Based on data from Language and Learning Foundation scaling up Structured Pedagogy in India

More resources and tools coming soon to a
theater near you

Zooming back out

Small shifts can make a big difference

Measurement, leaps from efficacy trials, etc.



This room is filled with the key actors in global education – researchers, policymakers, practitioners, funders.

Let's exchange and learn together.