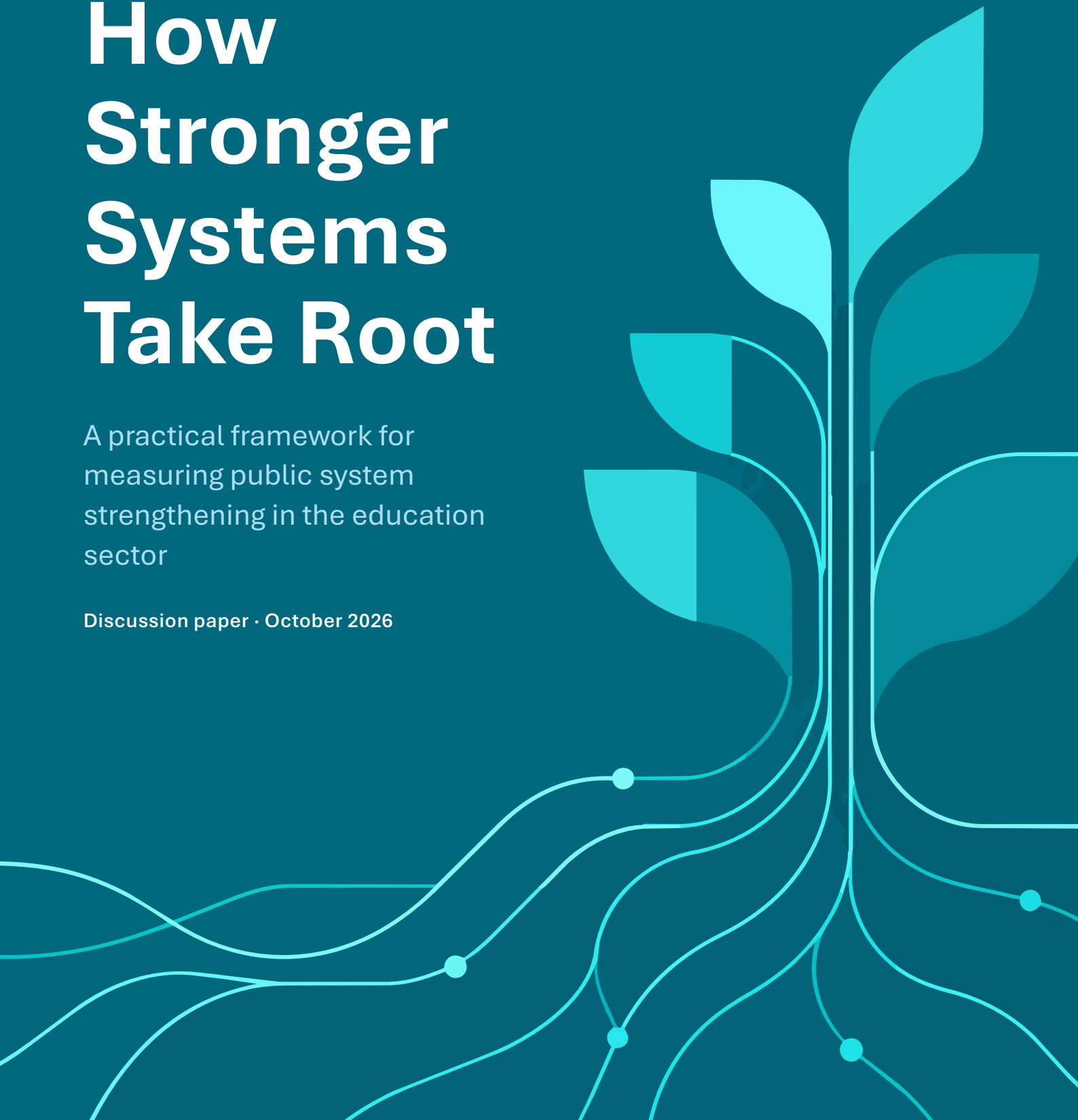


# How Stronger Systems Take Root

A practical framework for  
measuring public system  
strengthening in the education  
sector

Discussion paper • October 2026



## TABLE OF CONTENTS

|                                                                              |    |
|------------------------------------------------------------------------------|----|
| The core proposition                                                         | 3  |
| Context                                                                      | 3  |
| Introduction: The Elimu-Soko Measurement Framework for Systems Strengthening | 4  |
| How the evidence builds                                                      | 5  |
| Why a capability has to be inferred                                          | 6  |
| Before the cycle: is this a government priority?                             | 7  |
| Level 1: Has government taken the reform on?                                 | 8  |
| Condition 1: An owner inside government                                      | 9  |
| Condition 2: Money that recurs                                               | 10 |
| Condition 3: People who can do the work                                      | 11 |
| Condition 4: Evidence government steers by                                   | 11 |
| The parallel track: outcomes at government scale                             | 12 |
| How a cycle progresses, and what to expect when                              | 13 |
| Level 2: Does the reform survive the partner's exit?                         | 14 |
| Level 3: Can government do it again?                                         | 15 |
| Recording a cycle                                                            | 16 |
| Conclusion                                                                   | 20 |
| Sources and acknowledgements                                                 | 20 |

## The core proposition

A public system shows evidence of strengthening when it can take a proven intervention into routine delivery under its own authority. One intervention absorbed is evidence of institutional strengthening. An intervention still delivering benefits after intensive external support ends is stronger evidence of durability. A second cycle led by government for a problem it chose provides the strongest evidence of repeatable capability.

## Context

**Promising education innovations too often remain donor-funded projects:** they improve outcomes while support lasts but never become part of how government routinely delivers education. Tighter aid budgets make this longstanding problem more urgent. As described in [Embedding Private Sector Innovations in Government Systems](#) (Abdi and Vussonji, 2026), Elimu-Soko works with governments to select innovations that address their priorities, test and refine them through existing public institutions, and integrate effective approaches into recurrent budgets and routine delivery, with external funding supporting the transition.<sup>1</sup> This raises a practical question: **how do we know whether these innovation partnerships are strengthening the public system's own capability to deliver and sustain better outcomes?** This paper sets out a practical measurement framework for answering that question. We intend to test and refine it through our upcoming government engagements, and invite governments, funders and practitioners to use it and share what they learn.

“

Without meaningful ownership, even well-funded interventions fail to take root in national systems or survive beyond donor timelines.

”<sup>2</sup>

1. Abdi KM, Vussonji D. Embedding Private Sector Innovations in Government Systems. Stanford Social Innovation Review; January 21, 2026.

2. Ibid.

# Introduction: The Elimu-Soko Measurement Framework for Systems Strengthening

**Health researchers made the distinction first.** Chee and colleagues separated strengthening a health system from supporting it. A round of training, a new tool or an extra delivery unit adds capacity, can raise outcomes while the funding lasts, and says little about whether the system can finance, staff, run and replace that capacity on its own.<sup>3</sup> A decade later the label still does that work: most strengthening claims rest on inputs and coverage counted during the grant, and the systems that have gone further are recorded as country stories because nobody built a test they could pass. Education has the same gap. The sector reads children reached, teachers trained and budgets announced as evidence the system is strengthening, and most proven interventions stall at the point where government would have to change its rules, money and roles to carry them.

## Three terms used throughout

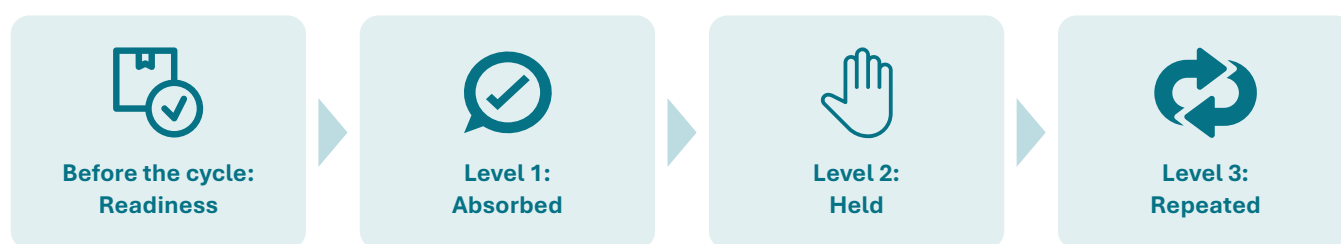
- **An intervention** is a practice or programme that can be piloted and evaluated, such as coaching for teachers or a structured reading programme.
- **A reform** is a change in how the public system works: its rules, budgets, roles and routines. It usually carries several interventions.
- **An investment** is what a funder puts in, whether to pilot an intervention, to move it into a reform, or both.

**The question is how you would know a system has been strengthened, as opposed to knowing that an intervention inside it worked.** The capability to adopt and sustain effective reforms cannot be observed directly; it is inferred from its exercise. The evidence comes in three levels. Government took one reform on, which four conditions show. The reform held after intensive support was withdrawn, and leadership changed, which a sustainment test shows. Government ran another cycle on its own initiative, which shows it can do it again.

**The field's current answer stops at the first level and declines to generalise even there.** Each investment is measured against its own theory of change. The evidence is specific to the intervention, the country and the moment, so judging it investment by investment makes sense. The conditions a public system needs to carry any reform without outside help are general. If nobody measures them, a set of absorbed interventions never adds up to a claim about a system, and a ministry or funder has nothing on which to compare one reform with another.

**These measures serve three decisions.** Inside a cycle, government decides whether to adopt, adapt or stop, and what evidence each choice needs. Across a cycle, a funder decides whether to continue, transition or exit, and when. Across cycles, government and its partners check whether government started the next cycle itself and needed less outside help than before. That comparison tests whether the capability demonstrated in one reform extends to another. Government also uses these measures to assess whether the partner has met its agreed commitments.

3. Chee G, Pielemeier N, Lion A, Connor C. Why differentiating between health system support and health system strengthening is needed. *International Journal of Health Planning and Management* 2013; 28(1): 85–94.



## How the evidence builds

| Level                              | The claim                                                                                                 | The evidence                                                                                                                                                                                 | When                                                     |
|------------------------------------|-----------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------|
| <b>Before the cycle: readiness</b> | The problem is a government priority, and the reform has a home, named decision-makers and a fiscal path. | The gate. A go or no-go before intensive piloting.                                                                                                                                           | Before piloting                                          |
| <b>Level 1: Absorbed</b>           | Government took the reform on.                                                                            | Four conditions, each read at the point of use, with outcomes holding under government-led delivery.                                                                                         | During the cycle; indicative pathway of years one to six |
| <b>Level 2: Held</b>               | The reform survived the withdrawal of intensive support, and a change of leadership.                      | The four conditions re-run 18 to 24 months after exit by a reviewer that government approves, with outcomes above the agreed minimum standard and any deterioration corrected by government. | After exit                                               |
| <b>Level 3: Repeated</b>           | Government can run the cycle itself.                                                                      | A second cycle led by government, from choosing the priority to financing and managing it, with less outside help than the first.                                                            | Cycles later                                             |

Most of the sector's effort sits at Level 1. Levels 2 and 3 reuse its measures later, applied by a reviewer that government approves. Level 1 shows the reform was absorbed; Level 2 tests whether it lasts; Level 3 shows whether government can do it again.

## Why a capability has to be inferred

**Ownership and motivation decide whether any condition is real.** A ministry can have the budget line, the trained staff and the data system and still not run the reform, because the people meant to run it feel no ownership of it. If no one senior is responsible for the reform, it drifts. A budget line no one will defend goes in the first squeeze. Ownership cannot be measured by asking, because the answer describes the relationship with the partner. It shows in what people do with a condition once it exists, and in what survives when support is withdrawn.

**So every condition is read at the point of use.** A budget can be allocated and never released. Staff can be trained and never deploy the training. A data system can run and change no decision. The condition counts when the money reached the school, the supervisor corrected the problem, the manager acted on the data. The “do not mistake for strengthening” column under each condition lists the presence-without-use cases that most often pass for absorption.

**Levels 2 and 3 test ownership under stress, where it is easiest to see.** Leadership changes during most cycles. If a reform survives its champion’s departure, others in government own it too. If it collapses, it depended on that one person. A second cycle, run for a problem government chose, shows that government values the capability and can reuse it.

**The unit of analysis is the public system.** It includes laws, mandates, budgets, staff, supervision, training, data and operating routines, and the political economy around them: who holds authority, who has reason to act, who can block. Measurement has to catch both formal institutionalisation and observable government behaviour, because a written commitment shows intent and behaviour shows practice.

## Before the cycle: is this a government priority?

### Are the conditions for adoption credible before a pilot begins?

**An intervention enters a pilot either as a response to a government priority or as a solution looking for a home, and the gate decides which.** The gate sits before the four conditions, and government and partner decide go or no-go together. Its first test is whether government would have worked on this problem without the partner. Government states its priorities before any intervention is discussed, and the test looks for time or money government has already spent on the problem. The others are named decision-makers, an explicit use for the evidence, and a first map of fiscal, institutional and political feasibility. The gate tests the partner too: whether its offer fits government's priority, budget and systems, or asks government to build around the intervention. An intervention that enters without passing these tests can succeed in the pilot and still have no route into government.

The gate is also the first, weak signal of Level 3. A cycle that begins from government's own priority is already a partial exercise of the capability Level 3 tests for, and the question the gate asks at the start of a cycle is the one a second government-led cycle answers at the end.

| What to measure                                                                                                                                                                                                                                                                                                                                                                                                                                       | What good looks like                                                                                                                                                                                                                                                                         | Do not mistake for strengthening                                                                                                                                                                                                                                                                                    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> <li>• Government-defined priority and problem statement</li> <li>• Named political and technical decision-makers</li> <li>• Explicit government use case for the evidence</li> <li>• Initial map of fiscal, institutional and political feasibility</li> <li>• Time or money government has already spent on the problem</li> <li>• Fit between the partner's offer and government's budget and systems</li> </ul> | <ul style="list-style-type: none"> <li>• The problem being tested is one government was already trying to solve</li> <li>• Decision-makers agree what evidence would trigger adopt, adapt or stop</li> <li>• A plausible adoption pathway exists before intensive piloting begins</li> </ul> | <ul style="list-style-type: none"> <li>• Generic government "buy-in"</li> <li>• A memorandum of understanding without decision authority</li> <li>• A donor-defined pilot looking for a government home after design</li> <li>• A priority named in a sector plan with no owner, time or money behind it</li> </ul> |



## Level 1: Has government taken the reform on?

**Once a pilot works, what must become true for the public system to carry it with no one from outside in the room?** Take the partner away and ask what breaks. Four things have to be in place or delivery stops.

- **An owner inside government.** The reform needs a permanent role in the ordinary machinery. A champion or a project unit is borrowed attention, and the reform ends when attention moves.
- **Money that recurs.** A reform that depends on a grant line or an annual battle stops when the grant ends. Absorption means the money flows through channels the system was going to fund anyway.
- **People who can do the work.** Training, supervision, quality assurance, data and procurement have to sit with people the system already employs and manages. If pulling the partner out breaks delivery, the function was rented.
- **Evidence government steers by.** Government needs information it uses itself to supervise, adapt and defend the reform when it is challenged. A monitoring system built for the donor report does not count.

Each is necessary: remove one and the reform cannot run unaided. Together they let the reform run tomorrow. Whether it keeps running after the partner leaves is Level 2's question.

# Condition 1:

## An owner inside government

Has government moved from endorsement to consequential decisions and normal operating practice?

**Ownership shows in behaviour.** The signal is the reform appearing in the ordinary machinery of government: workplans, staff roles, supervision routines, curriculum guidance, regulation, budget processes, data systems and management decisions. It holds only when it has depth below the senior champion, in the middle layer of officials and teachers who live the reform day to day and carry it out as part of their normal work. Name whose ownership matters: political and fiscal authorities who authorise and fund the reform; officials, school leaders and teachers who run it; and parents, civil society and professional groups whose support helps sustain it.

**Ownership and capability are different things**

This condition is about authority: who decides, who is accountable, whose job the reform becomes. Condition 3 is about capability: whether the system can perform the work with the partner off the critical path. A ministry can hold authority over a reform its staff cannot yet deliver, or deliver a reform nobody senior has taken responsibility for. Both have to become true, and they move at different speeds.

| What to measure                                                                                                                                                                                                                                                                                                                                                            | What good looks like                                                                                                                                                                                                                                                                                                               | Do not mistake for strengthening                                                                                                                                                                                                                      |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"><li>• Government-led problem definition and selection</li><li>• Formal adoption decisions with consequences</li><li>• Integration into policy, standards, curriculum or operating guidance</li><li>• Named officials and departments performing the work</li><li>• Routine supervision and management incorporating the reform</li></ul> | <ul style="list-style-type: none"><li>• Government moves from participant to decision-maker</li><li>• The reform is part of normal operating expectations, not a special project</li><li>• Middle-layer staff carry the reform as part of their normal work</li><li>• Ownership survives changes in individual champions</li></ul> | <ul style="list-style-type: none"><li>• Endorsement, steering committees or meeting attendance</li><li>• A policy document that is not implemented</li><li>• Partner activity inside government buildings presented as government ownership</li></ul> |

## Condition 2:

### Money that recurs

Can the reform operate within the fiscal envelope and durable financing of the system?

**Impact without fiscal fit is incomplete.** Fit means the recurrent cost sits inside budgets government already funds, and the cost question has to be answered during design and piloting, before a successful pilot has created expectations of scale it cannot fund. Fit early in the cycle shows whether government is likely to carry the cost later.

**Cost informs the decision**

Where the reform is unaffordable at current costs, government and partner redesign the cost structure or reconsider the pathway. Costs move: unit prices fall, delivery models change, a government reprioritises. A student-facing tool whose per-user cost the recurrent budget cannot carry needs a different cost structure before absorption is realistic; the adopt, adapt or stop decision stays with government.

| What to measure                                                                                                                                                                                                                                                                                                                                                                                              | What good looks like                                                                                                                                                                                                                                                                                                              | Do not mistake for strengthening                                                                                                                                                                                                                                                   |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"><li>• Cost per learner, teacher or service unit at pilot and projected government scale</li><li>• One-time transition costs separated from recurrent delivery costs</li><li>• Government resources already available: personnel, institutions, infrastructure, budget lines</li><li>• Public allocation, release, procurement, expenditure and frontline receipt</li></ul> | <ul style="list-style-type: none"><li>• Cost trajectory moves toward government price points during the pilot</li><li>• External financing covers temporary transition infrastructure, not permanent recurrent costs</li><li>• At steady state, recurrent costs are absorbed by the durable system that owns the reform</li></ul> | <ul style="list-style-type: none"><li>• Large external grants as evidence of sustainability</li><li>• A budget allocation that is never released or spent</li><li>• Co-financing treated as the destination while the model still depends structurally on external money</li></ul> |

**Money moves from allocation through release, procurement and expenditure to frontline delivery.** A budget line is evidence of intent. Absorption requires following the money through to the service, and many reforms that look financed break between allocation and release. Funders matter here too: pilot funding, transition funding and steady-state public financing solve different problems, and funding only the pilot builds a cliff into the design.

## Condition 3:

### People who can do the work

**Can government staff and permanent institutions deliver, supervise and manage the reform without a parallel project structure?**

**Workforce matters, but workforce alone is too narrow.** What counts is whether the delivery system can perform the functions an external implementer used to carry: training, supervision, management, quality assurance, data, procurement, technology, materials and adaptation.

| What to measure                                                                                                                                                                                                                                                                                                                                        | What good looks like                                                                                                                                                                                                                                                                                                                                                           | Do not mistake for strengthening                                                                                                                                                                                                                        |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> <li>Government personnel delivering the model</li> <li>Pre-service and in-service training embedded in permanent institutions</li> <li>Supervision, coaching and quality assurance through existing structures</li> <li>Management, procurement, data and operational routines transferred to government</li> </ul> | <ul style="list-style-type: none"> <li>The external role moves from leading, to joint delivery, to government leading with partner support, to government operating alone</li> <li>The middle layer of district officials, supervisors, training institutions and school leaders carries the reform</li> <li>The external partner is no longer on the critical path</li> </ul> | <ul style="list-style-type: none"> <li>People trained without retention or use</li> <li>A train-the-trainer event with no institutional home</li> <li>Government staff present while external personnel still perform the critical functions</li> </ul> |

**Track which functions named in the absorption hypothesis,** the plan written before the pilot, still depend on the external partner, and whether any is indispensable to routine delivery. A healthy pathway shows direct delivery and operational dependency declining while government responsibility rises. The partner may stay valuable as evidence partner, translator, broker or technical adviser. It should not stay indispensable to routine delivery. Government may decide to keep buying a function from outside, on its own budget. That is government's decision, and the function counts as government's.

## Condition 4:

### Evidence government steers by

**Can the system generate and use decision-ready evidence to supervise, adapt and defend the reform?**

**The evidence has to serve government's decisions.** Government needs evidence to decide, budget, supervise, adapt and defend a reform, and the condition is whether the system produces that evidence and uses it. Data infrastructure counts when it changes decisions.

| What to measure                                                                                                                                                                                                                                             | What good looks like                                                                                                                                                                                                                                                                                                                     | Do not mistake for strengthening                                                                                                                                                                                                                                     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> <li>Government-owned data and monitoring systems</li> <li>Common metrics and population benchmarks</li> <li>Implementation quality during scale-up</li> <li>Cost, delivery and outcome data reviewed together</li> </ul> | <ul style="list-style-type: none"> <li>Officials can answer: adopt, adapt or stop?</li> <li>Middle-layer managers use data to correct implementation problems</li> <li>The system detects quality deterioration as external support declines</li> <li>Measurement infrastructure stays functional after individual grants end</li> </ul> | <ul style="list-style-type: none"> <li>Parallel dashboards built only for donor reporting</li> <li>Impact evaluations used as a substitute for routine management information</li> <li>Large volumes of data with no decisions that changed because of it</li> </ul> |

# The parallel track: outcomes at government scale

Do outcomes hold, and improve, when the system owns and runs the reform?

**Absorption is worth nothing if outcomes do not hold.** The four conditions measure whether the system has taken the reform on; outcomes measure whether that was worth doing. The two run on different clocks, and reading either alone misleads. System capability can deepen before outcomes move, and usually does. Outcomes can improve without the system changing, when a well-supported pilot lifts results while the system stays externally dependent. Outcomes can also dip during transition while absorption deepens, as delivery moves from partner to government and quality wobbles before it recovers. So measure both and expect them to diverge. An investment on track can show deepening absorption and flat outcomes for a period, or strong outcomes and shallow absorption that will not survive the partner leaving.

| What to measure                                                                                                                                                                                                                                                                          | What good looks like                                                                                                                                                                                                                                           | Do not mistake for strengthening                                                                                                                                                                                                                                         |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"><li>• Population reach and coverage</li><li>• Equity of access and outcomes</li><li>• Service or instructional quality</li><li>• Learning, development or other intended outcomes</li><li>• Cost-effectiveness under government-led delivery</li></ul> | <ul style="list-style-type: none"><li>• Pilot efficacy survives the move to government-led effectiveness</li><li>• Quality stays above the agreed minimum standard as scale expands</li><li>• Outcomes are achieved at a cost the system can sustain</li></ul> | <ul style="list-style-type: none"><li>• Beneficiaries reached as a stand-alone systems metric</li><li>• Pilot results generated under unusually intensive external support</li><li>• Scale counted as success while quality or outcomes materially deteriorate</li></ul> |

## How a cycle progresses, and what to expect when

**Each condition moves through stages, and the conditions move at different speeds.** A stage applies to each condition separately: ownership can be government-led while delivery is still shared. The aim is a credible trajectory from external dependence toward government ownership, and government and funders read the trajectory, not a score.

| External                                                                                                                    | Shared                                                                      | Government-led                                                                                 | Embedded                                                                                                                            |
|-----------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|
| Priority, design and delivery are largely external. Government participates. Clearing the gate is the exit from this stage. | Government and partner share delivery, supervision, evidence and financing. | Government leads routine delivery and management. The external role is targeted and declining. | The reform is normal public-system operation, recurrent costs and core functions are government-held, and Level 2 has confirmed it. |

**Funders need to know what to expect, and when.** The sequence below illustrates a cycle of roughly six years; its first three periods run through year six. These are planning assumptions, not deadlines or universal estimates. A funder seeking absorption should reserve funding for the transition and agree when evidence under government-led delivery will be sufficient to judge results. Most grants close before that point, which is why the sector has so many evaluated pilots and so few judged reforms.

| Period                            | What absorption should show                                                                                                                      | What outcomes should show                                                                                                                                                    |
|-----------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Years one to two</b>           | The gate is cleared, a government owner is named, the absorption hypothesis is written and the cost at scale is known.                           | Pilot outcomes under external delivery, labelled as such.                                                                                                                    |
| <b>Years two to four</b>          | Delivery is shared. Money is moving through the chain, at least to release. Functions are transferring and dependence on the partner is falling. | Outcomes may remain flat or dip during transition. Assess any deterioration against the agreed minimum standard and act on its cause; a dip is not automatically acceptable. |
| <b>Years four to six</b>          | Government leads routine delivery and management. The external role is targeted and declining.                                                   | Outcomes under government conditions become the primary measure, and the question shifts to whether the reform works at the price and quality the system can hold.           |
| <b>18 to 24 months after exit</b> | Level 2: a reviewer that government approves re-runs the four conditions. What survives is the verdict on absorption.                            | Read against both the agreed minimum standard and the pre-exit level, with regression named as regression.                                                                   |



## Level 2: Does the reform survive the partner's exit?

**The four conditions describe where a reform is. This level asks whether it stays there when nobody is holding it.**

Absorption is not linear and regression is a normal feature of systems. The question is whether absorption is deepening or reversing as support recedes and leadership turns. These signs show it is reversing:

- A government champion leaves and no replacement takes ownership.
- A budget is allocated but not released.
- Quality declines as government delivery expands.
- A data system stops functioning or reverts to a parallel project structure.
- Frontline staff treat the reform as additional work rather than normal work.
- External partners resume functions government had previously absorbed.

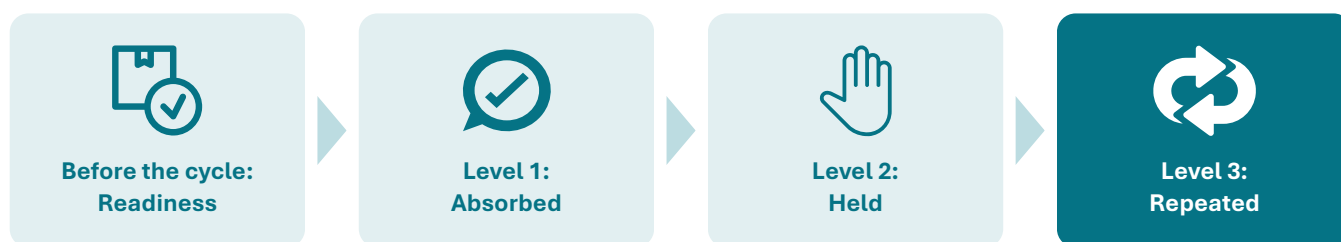
**Every progress indicator has a regression twin, and the two are recorded together.** If a budget allocation is progress, non-release is regression. If government-led training is progress, a return to externally delivered training is regression. The pairing keeps the picture honest during the cycle, and it gives the sustainment test something to compare against.

### The sustainment test

18 to 24 months after intensive support has reduced or ended, a reviewer that government commissions or approves re-runs the four conditions and asks two further questions. Do outcomes stay above the agreed minimum standard, at a cost the recurrent budget carries? Where quality or delivery deteriorated, did government detect and correct it without the partner? A reform that still runs, but where government neither noticed nor fixed a decline, is recorded as “fragile”. What survives is more informative than what existed while the grant was live, and it is the clearest evidence available of whether ownership has held. A system not yet tested this way is recorded as “not yet tested”.

**Durability is political as well as technical.** Reforms endure when institutions, leaders and constituencies have an interest in keeping them alive: communities, professional groups, local government, civil society, parent structures, frontline workers, political champions. The political and social environment shifts as a reform matures, so this is assessed at every stage, and the verdict is read here.<sup>4</sup> Four questions carry it. Who benefits from the reform and can say so? Who has an institutional incentive to defend it when budgets tighten or leadership changes? Can users and frontline actors influence the system? Would reversal be visible to constituencies that matter politically or administratively? The test distinguishes durable commitment from political convenience that looks like it. A reform adopted for opportunity can show every formal marker of ownership and fade the moment the opportunity passes.

4. On sustainability as a process that runs from the start of a programme, see Pluye P, Potvin L, Denis JL. Making public health programs last: conceptualizing sustainability. *Evaluation and Program Planning* 2004; 27(2): 121–133.



## Level 3: Can government do it again?

**A system that has absorbed one reform and held it has been strengthened for that reform.** A second government-led cycle is the strongest evidence that the capability extends beyond the first. The test is whether government can put the four conditions in place again for a problem it chose, under its own authority and with financing it can sustain. Condition 1 measures ownership of a reform; only repetition shows ownership of the practice of adopting.

**A second cycle counts when government leads every step.** It chose the problem. It ran its own gate, naming decision-makers and a fiscal path before piloting. It financed the pilot and the transition from its own envelope, or raised the money on its own terms. It managed the cycle with the partner absent or advising. The evidence is the cycle itself, recorded on the same page as the first and compared with it: whether government needed less outside help the second time, with fewer functions performed by the partner, more of the money from its own budget, and less time from approval to frontline delivery, read against the scale of the second reform.

Few reforms have been followed long enough to produce this evidence. Elimu-Soko's oldest work, in Rwanda and Zanzibar, is one cycle each and only now old enough for the sustainment test. That is where most of the field stands, and why the measures here concentrate on Level 1: absorption has to be solved first, and it is where the evidence and the money currently fall short. In both, government adopted the reform and now funds it. These are claims of adoption and financing; durability remains to be tested, and repeatable capability has not yet been shown.

## Recording a cycle

Every cycle is recorded on the same page, so a portfolio can be compared by trajectory without a common indicator set. The page makes an adoption conversation explicit and gives government and funders one legible picture. The partner rates each condition by stage from behavioural evidence, records the evidence behind the rating, and treats the evidence note as the substance and the mark as the summary. Government rates the partner against the commitments in the absorption hypothesis. The hypothesis can change at a review. Each change is written beside the original, with the reason and who asked for it, and the changes are read together at every review. An intervention that keeps needing adaptation to be absorbed is telling government something about fit, and the review treats repeated adaptation as evidence for stopping. Government and partner sign the page, and where they disagree, both views are recorded. A reviewer that government approves completes the Level 2 and Level 3 rows. Government and funders look for rightward movement across cycles and for evidence notes that name the next decision.

**Between reviews, six signals show the direction of travel.** There is one for the gate, one for each condition and one for outcomes, and each has a regression twin read alongside it.



| Element                                   | Leading indicator                                                                                                                                                            | Regression twin                                                                                 |
|-------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| <b>Gate: readiness</b>                    | Officials name the problem when they list their own priorities, before anyone mentions the intervention.                                                                     | Officials raise the problem only when asked about the intervention.                             |
| <b>1. An owner inside government</b>      | At the last breakdown, a government official noticed and acted before the partner did. A period with no breakdown is recorded as not tested.                                 | The partner fixed the breakdown and told government afterwards.                                 |
| <b>2. Money that recurs</b>               | Early in the cycle, unit cost falls toward what government can pay. Later, government releases a rising share of the recurrent cost and the money reaches the frontline.     | Unit cost stops falling, or government allocates money that is never released or never arrives. |
| <b>3. People who can do the work</b>      | At each review, government independently performs more of the essential functions, with no indispensable delivery function left dependent on the partner at full absorption. | The partner takes back a function government had taken on.                                      |
| <b>4. Evidence government steers by</b>   | Government changed a delivery decision on its own data.                                                                                                                      | Monitoring reverts to reporting for the funder.                                                 |
| <b>Outcomes under government delivery</b> | Implementation quality under government delivery stays above the agreed minimum standard.                                                                                    | Coverage grows while quality or outcomes fall.                                                  |

| Element                                          | Status                                                                                                                                                                                                                               | Evidence / next decision |
|--------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|
| <b>Gate: readiness</b>                           | <input type="radio"/> Not cleared <input type="radio"/> Cleared                                                                                                                                                                      |                          |
| <b>1. An owner inside government</b>             | <input type="radio"/> External <input type="radio"/> Shared <input type="radio"/> Gov-led<br><input type="radio"/> Embedded                                                                                                          |                          |
| <b>2. Money that recurs</b>                      | <input type="radio"/> External <input type="radio"/> Shared <input type="radio"/> Gov-led<br><input type="radio"/> Embedded                                                                                                          |                          |
| <b>3. People who can do the work</b>             | <input type="radio"/> External <input type="radio"/> Shared <input type="radio"/> Gov-led<br><input type="radio"/> Embedded                                                                                                          |                          |
| <b>4. Evidence government steers by</b>          | <input type="radio"/> External <input type="radio"/> Shared <input type="radio"/> Gov-led<br><input type="radio"/> Embedded                                                                                                          |                          |
| <b>Outcomes under government delivery</b>        | Trend: <input type="radio"/> Improving <input type="radio"/> Holding<br><input type="radio"/> Deteriorating.<br>Minimum standard: <input type="radio"/> Met <input type="radio"/> Not met<br><input type="radio"/> Not yet assessed. |                          |
| <b>Level 2: sustainment</b>                      | <input type="radio"/> Not yet tested <input type="radio"/> Holding <input type="radio"/> Fragile<br><input type="radio"/> Regressing                                                                                                 |                          |
| <b>Level 3: second cycle</b>                     | <input type="radio"/> Not started <input type="radio"/> Started <input type="radio"/> Needed less help                                                                                                                               |                          |
| <b>Partner obligations (rated by government)</b> | <input type="radio"/> Met <input type="radio"/> Partly met <input type="radio"/> Not met                                                                                                                                             |                          |
| <b>Rated by</b>                                  | Partner: _____<br>Government: _____<br>Reviewer (Levels 2 and 3): _____                                                                                                                                                              |                          |

### Illustrative example

Hypothetical teacher-coaching reform illustrates how to use the framework:

Government releases recurrent funding for coaching, district supervisors run school visits, and learning outcomes remain above the agreed minimum standard. However, the partner still analyses delivery problems and organises the response; district staff do not yet perform this function independently.

The judgement is partial absorption. Financing and routine delivery are government-led, but diagnostic and corrective capability remains shared and the partner is still on the critical path. The next decision is to strengthen government's use of its own data and demonstrate independent problem-solving before further expansion. Sustainment is not yet tested and a second cycle has not started.

### What the partner owes

At every review, the government counterpart answers three questions about the partner, against the commitments in the absorption hypothesis:

- Did the partner step back from each function on the date it committed to, and stay stepped back?
- Did it build data for the ministry's decisions or for the funder's report?
- Did the reform cost what the partner projected?

**The page is only as good as the cycle it records, so the cycle is designed to be measurable before it starts.** Every pilot answers one question: adopt, adapt or stop. The decision-maker, the decision date and the evidence required are named before the pilot begins. The absorption hypothesis is written down and answers six questions:

- Which government unit owns the reform?
- Which budget or existing resource base sustains it?
- Which people and institutions do the work?
- Which data systems manage it?
- Which functions does the partner perform now, and on what date does each pass to government?
- What will the reform cost per unit at scale?

Learning outcomes alone are never enough; the evidence has to cover cost at scale, delivery feasibility, readiness and what the transition demands.

**Many conventional metrics stay useful. They need labelling correctly.** The following are programme or transition indicators, and none is evidence at any of the three levels unless it is tied to deeper absorption:

- |                                              |                                                                                           |
|----------------------------------------------|-------------------------------------------------------------------------------------------|
| • children reached during the grant;         | • officials attending steering committees;                                                |
| • teachers or frontline workers trained;     | • a budget line announced but not executed;                                               |
| • memoranda of understanding signed;         | • a pilot shown to work under intensive external delivery;                                |
| • policies, standards or curricula produced; | • a programme at national scale while recurrent delivery stays structurally donor-funded. |

### The discipline

Ask what the system absorbed, what it can now do without indispensable partner support, and whether it has exercised that capability again. Let the claim be no larger than the evidence, and apply the rule in both directions: absence of evidence is not evidence of absence.

## Conclusion

**The sector counts activities and outcomes inside projects, then infers that the public system strengthened.** A better test asks three questions in order and leaves the third until the first two are answered. Did the system take the reform on: is there an owner inside it, money that recurs, people who can run it, evidence it steers by, and outcomes that hold under its own delivery? Did it sustain the reform after intensive external support ended and leadership changed? And has it exercised the capability again, for a problem it chose, on its own initiative?

## Sources and acknowledgements

The context of this paper is based on [“Embedding Private Sector Innovations in Government Systems”](#) (Abdi and Vussonji, 2026).

The four conditions are derived from first principles about how a public system carries a reform. They draw on three bodies of work:

- the health-systems distinction between supporting a system and strengthening it (Chee and colleagues, 2013);
- the programme-sustainability literature on institutionalisation, maintained benefits and community capacity (Shediac-Rizkallah and Bone, 1998; Pluye and colleagues, 2004);<sup>5</sup>
- research on how new practices become routine in public bureaucracies (Yin, 1979).<sup>6</sup>

This framework was developed by Alim Ladha, Devang Vussonji, Dr Muna Ngenda, and Keshinee Shah.

5. Shediac-Rizkallah MC, Bone LR. Planning for the sustainability of community-based health programs: conceptual frameworks and future directions for research, practice and policy. *Health Education Research* 1998; 13(1): 87–108.

6. Yin RK. *Changing Urban Bureaucracies: How New Practices Become Routinized*. Lexington Books, 1979.

## How to cite

Ladha, A., Vussonji, D., Ngenda, M. and Shah, K. (2026). How Stronger Systems Take Root: A practical framework for measuring public system strengthening in the education sector. Elimu-Soko Discussion Paper, October 2026.

