
Measuring what matters

An evidence framework for sustainability
education

The sector reports participation and calls it impact. This paper sets out five levels of proof, an indicator library a programme can actually run, and three rules that keep the numbers honest.

The argument in one page

Sustainability education is mostly reported in units that cannot fail: participants reached, certificates issued, awareness raised. Every one of those numbers can grow while nothing in a classroom changes. They measure the size of the effort, not the existence of the effect.

The field knows how to do better, because its best evidence already looks different. The global best-buys ranking is built on measured learning outcomes, not attendance.¹ The trial that moved this series measured achievement in standard deviations against cost.² And the most expensive lesson in the sector, \$18,000 per teacher per year with no consistent improvement, was only visible because somebody measured practice instead of attendance.³

What programmes lack is not rigour in principle but a framework they can run without a research department: what to count, with which instrument, owned by whom, at which level of proof.

This paper proposes that framework: five levels, from participation to institutional change, each with indicators a working programme can collect, and three rules that keep the numbers honest. We run it on our own programmes first, and publish the misses with the scores.

FIGURE 1 • WHERE THE REPORTING STOPS



¹ GEEAP, SMART BUYS 2020, UPDATED 2023 • ² JACKSON & MAKARIN 2018, AEJ: ECONOMIC POLICY • ³ TNTP 2015, THE MIRAGE

01 What gets counted, and what decisions need

The gap is not ignorance. It is convenience. Participation is counted at the door, awareness is a survey, and both are available the same week the programme ends. Learning takes an instrument, behaviour takes a follow-up, and institutional change takes a year. Reporting gravitates to what is cheap, and the sector's annual reports show it.

The cost of convenient numbers is misallocation. A funder comparing two programmes by reach will fund the louder one. The same funder comparing artefacts of learning, checks passed, projects delivered, funds the one that works. Evidence standards are not paperwork. They are how money finds the thing that teaches.

FIGURE 2 • WHAT REPORTS CARRY, WHAT DECISIONS NEED

CARRIED IN ANNUAL REPORTS



available the week the programme ends, and unable to fail

NEEDED FOR A FUNDING DECISION



dashed because mostly absent from the reporting, not because unmeasurable

CONCEPTUAL • DASHED MEANS ABSENT FROM TYPICAL REPORTING

02 Five levels of proof

The levels are cumulative. Each one is legitimate to report, provided it is named as what it is. The failure mode is not reporting level one. It is reporting level one in the language of level five.

L1	Participation	Who showed up, for what, for how long. Necessary, cheap, and never sufficient. The floor, stated plainly.
L2	Learning outcome	An artefact or check produced inside the session: a measurement logged, a five-minute check passed. Designed into the resource, so collection costs the teacher nothing extra.
L3	Applied project	The learning used beyond the lesson: a proposal, a build, a measured change at home or school. The level where knowledge becomes behaviour, and the first level worth a site visit.
L4	Teacher capability	The teacher delivers again, unaided, and adapts: redelivery without support, reuse across classes, requests for the next unit. Capability is the multiplier every other level rides on.
L5	Institutional change	The system moves: the unit enters the timetable, the registry shows the band, the budget line recurs. The only level at which a programme can honestly claim to have changed education.

The evidence framework

Worked with live indicators from the NTRP kit Water from Air, so no cell is hypothetical.

LEVEL	INDICATOR • WHAT IS COUNTED	INSTRUMENT • HOW IT IS CAUGHT	OWNER
L1 Participation	classes taught with the kit, students present	library log, one line per delivery	the centre
L2 Learning	millilitres logged per workstation; five-minute check scores	the record card and quiz inside the kit	the teacher
L3 Application	capture proposals submitted for the school site	the cut-and-fold proposal card, photographed	the school
L4 Capability	redelivery without support; next-unit requests	library reuse data, support tickets	the centre
L5 Institution	units timetabled; bands shown on the registry; recurring budget lines	timetable and registry audit, annual	the authority

 RUNNING TODAY • L1, L2, L3
  PARTIAL • L4, BETA LIBRARY
  NOT YET • L5

03 Three rules of evidence

Name the source, every time. A number without a named instrument is a claim. Internal figures are labelled internal, and the label is not a demotion. It is what makes the number citable.

Modelled is never outcome. Projections, targets and reach estimates may be published, dashed and labelled, beside evidence, never as it. The 300-unit coverage target in Paper 05 is drawn as a dashed outline for exactly this reason.

Publish the misses. Paper 01 scored our own flagship kit and published the criterion it fails. This paper publishes that L5 is not yet met anywhere in our register. An evidence framework that only ever produces good news is a marketing framework.

For funders, CSR leaders and evaluators, the ask is one line in the grant template: state the level, name the instrument, report the misses. Programmes will meet the standard the money asks for, exactly as makers build to what the shelf rewards.

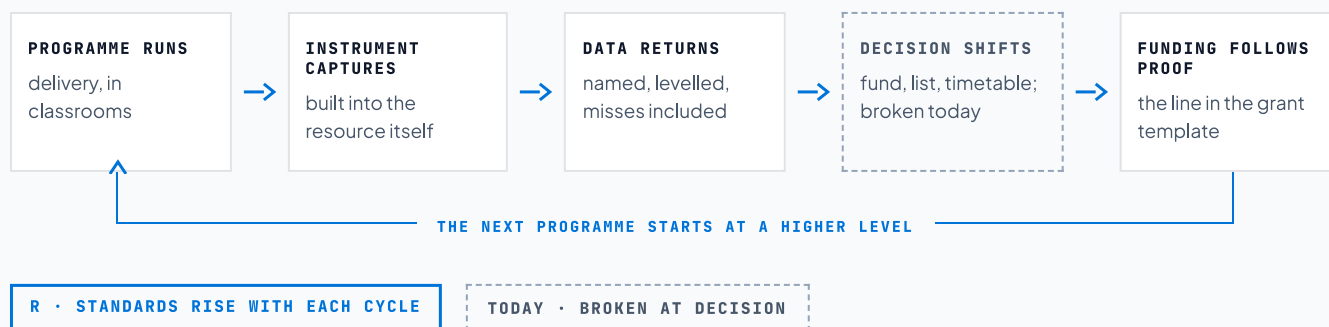
THE GRANT-TEMPLATE LINE

STATE THE LEVEL • NAME THE INSTRUMENT • REPORT THE MISSES

Count lessons taught, not licences issued.

The loop that money closes

Evidence only matters where a decision consumes it. The broken link is not collection. It is use.



THE SYSTEM IN NUMBERS

70%

of ten-year-olds in low- and middle-income countries in learning poverty, up from 57% before the pandemic

WORLD BANK ET AL. 2022

\$21 tn

potential lifetime earnings at risk for this generation of students

WORLD BANK ET AL. 2022

+0.08 σ

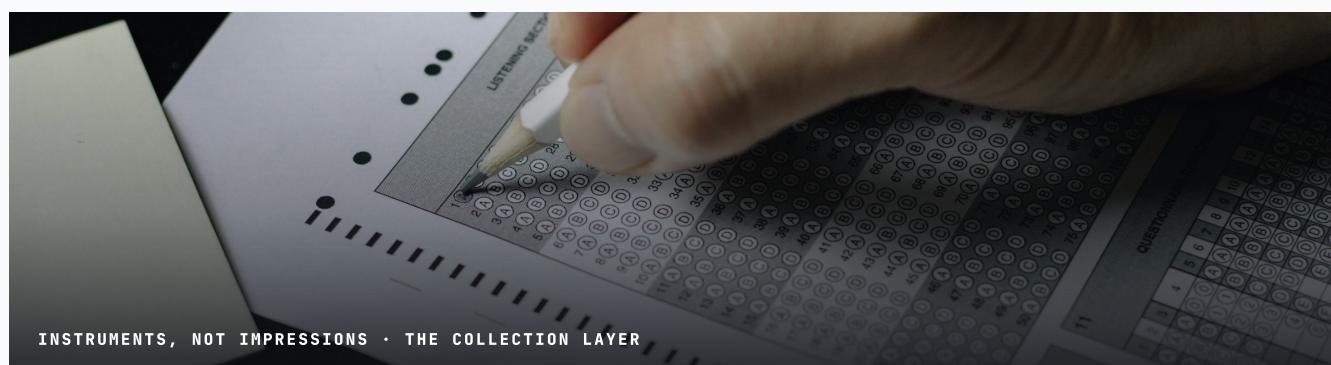
for \$430 per teacher: what measured evidence found when somebody looked

JACKSON & MAKARIN 2018

About this series

The Last Mile Papers are published by The Green Box (GBOX), a sustainability capability-building organisation founded at Cambridge Judge Business School in 2018, working across 16 countries with delivery hubs in Cambridge, Riyadh, Al Khobar, and Lahore. The series argues one idea from six directions: education priorities only create impact when teachers have practical resources, contextual capability, and measurable outcomes. Paper 06 closes the series with the evidence framework the other five papers are held to.

[GREENBOX.WORLD/NTRP](https://greenbox.world/ntrp)



Sources

- 01 Global Education Evidence Advisory Panel (2020; updated 2023). Smart Buys: Cost-Effective Approaches to Improve Global Learning. World Bank, FCDO, UNICEF.
- 02 Jackson, C. K., Makarin, A. (2018). Can Online Off-the-Shelf Lessons Improve Student Outcomes? American Economic Journal: Economic Policy, 10(3).
- 03 TNTP (2015). The Mirage: Confronting the Hard Truth About Our Quest for Teacher Development.
- 04 Darling-Hammond, L., Hyler, M. E., Gardner, M. (2017). Effective Teacher Professional Development. Learning Policy Institute.
- 05 GBOX (2026). NTRP library instruments: record cards, quick checks, proposal cards, library logs.
- 06 The Green Box (2026). The Last Mile Papers 01 and 05.
- 07 World Bank, UNESCO, UNICEF, FCDO, USAID, Gates Foundation (2022). The State of Global Learning Poverty: 2022 Update.

Positions, stated and revisable

The Last Mile Papers are opinion pieces. They present the positions The Green Box holds at the date on the cover, formed from the sources named in each paper and argued so they can be checked.

Positions can change. Where new evidence, corrected data, or a better argument arrives, the position moves and a revised edition says so. Later editions supersede earlier ones; the edition date on the cover is the record.

Internal figures are labelled internal. Modelled figures are labelled modelled and are never presented as evidence of outcome. Nothing in this series constitutes professional, legal, or investment advice.

© THE GREEN BOX 2026 • ALL RIGHTS RESERVED

Quotation with attribution is welcome. For reuse beyond quotation, write to hello@thegbox.org.

Awareness was never the output.

The series closes where it started: at the lesson, with
the evidence attached.

THE LAST MILE PAPERS • 01 TO 06 • GREENBOX.WORLD/NTRP

