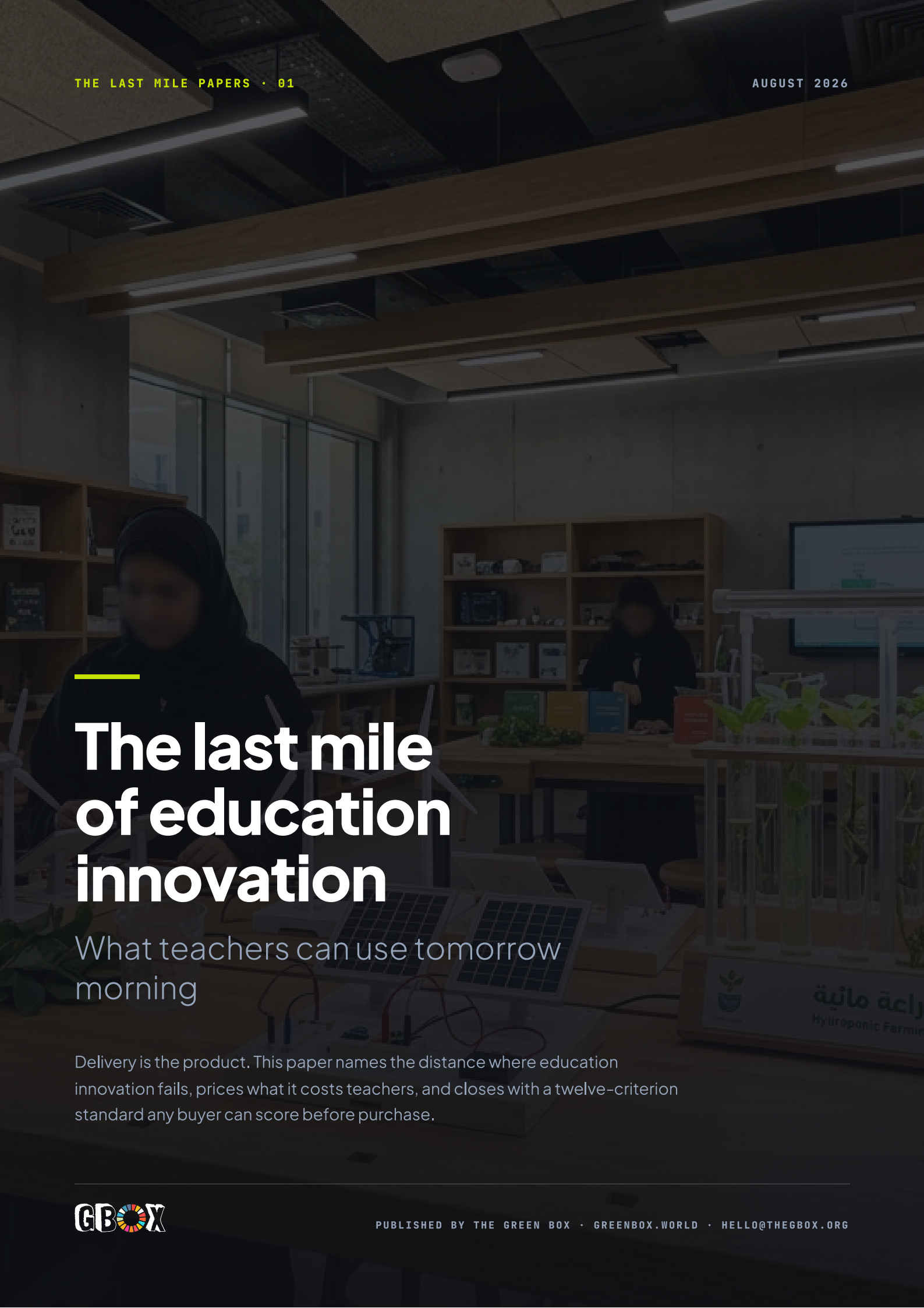
AUGUST 2026



The last mile of education innovation

What teachers can use tomorrow morning

Delivery is the product. This paper names the distance where education innovation fails, prices what it costs teachers, and closes with a twelve-criterion standard any buyer can score before purchase.

PUBLISHED BY THE GREEN BOX · GREENBOX.WORLD · HELLO@THEGBOX.ORG

The argument in one page

Education systems have never spent more on improving teaching, and teachers have never had less they can actually use. Large United States districts spend around $18,000 per teacher per year on professional development with no consistent link to improvement.[1] Two thirds of purchased education software licences go unused.[2] One Laptop per Child put roughly three million machines into classrooms and left no measurable learning gains behind in the most rigorous evaluation.[3]

These failures are usually filed under implementation, as if delivery were a detail. This paper argues the opposite: delivery is the product. Education innovation fails at the last mile, the distance between what a system buys and what a teacher can deploy in tomorrow's lesson with tonight's preparation time and the school's actual equipment.

The unit of adoption is not the school, the platform, or even the teacher. It is the lesson. Resources that do not survive contact with tomorrow morning do not exist, whatever the procurement record says.

The paper closes with a practical instrument: the Classroom-Readiness Standard, twelve criteria any funder, ministry, or maker can score from the resource itself in under an hour. We apply it to our own materials first, publish the results, gaps included, and invite the field to do the same.

FIGURE 1 · THE LAST MILE, DRAWN

1 TNTP 2015, THE MIRAGE · 2 GLIMPSE K12 2019 · 3 CRISTIA ET AL. 2012, INTER-AMERICAN DEVELOPMENT BANK, PERU

01 The graveyard of good ideas

In 2012, the Inter-American Development Bank published the first randomised evaluation of One Laptop per Child in Peru, the programme's largest deployment. Access to computers rose dramatically. Scores in mathematics and language did not move.[3] The laptops were real, the classrooms were real, and the distance between them swallowed a billion-dollar idea.

The pattern is not confined to hardware. TNTP's study The Mirage[1] followed more than 10,000 teachers across large American districts and found spending of roughly $18,000 per teacher per year, and about ten school days of teacher time, on development activities that produced no consistent improvement in practice. A Glimpse K12 audit of $2 billion in school spending found 67 percent of software licences never activated.[2] LearnPlatform's usage data shows a typical American district touching well over a thousand distinct digital tools in a single year,[4] a number that describes churn, not adoption.

Three different products, one failure mode. Each was optimised for the person who signs, not the person who teaches. The buyer sits in a ministry, a board meeting, or a conference hall, and is persuaded by vision, scale, and dashboards. The user stands in front of 32 students at 7:45 in the morning and is persuaded by exactly one thing: whether this works in the next 40 minutes.

Logistics companies solved their version of this problem by naming it. The last mile is the final leg of delivery, the shortest distance and the largest share of cost, where value either arrives or does not. Education has a last mile too. It runs from the resource to the lesson, and almost nothing in the sector is engineered for it.

1 TNTP 2015, THE MIRAGE · **2** GLIMPSE K12 2019, $2BN SPEND AUDIT · **3** CRISTIA ET AL. 2012, IDB, PERU · **4** LEARNPLATFORM BY INSTRUCTURE 2022

FIGURE 2 · THE WASTE LEDGER

$18,000

per teacher, per year, on development

NO CONSISTENT IMPROVEMENT FOUND

67%

of purchased software licences never activated

MAXIMUM · 100%

3,000,000

laptops shipped

MEASURABLE LEARNING GAINS: NONE

SOURCES: TNTP 2015 · GLIMPSE K12 2019 · CRISTIA ET AL., IDB 2012

02 The unit of adoption is the lesson

Consider the arithmetic of a teacher's week. Across OECD systems, lower-secondary teachers report about 38.8 hours of weekly work, of which roughly 6.8 hours is individual planning and preparation.[5] Spread across a typical teaching load, that is on the order of fifteen to twenty minutes of preparation per lesson taught. That quarter hour is the entire budget a new resource competes for. Anything that demands more is not competing with other resources. It is competing with marking, with family, and with sleep, and it loses.

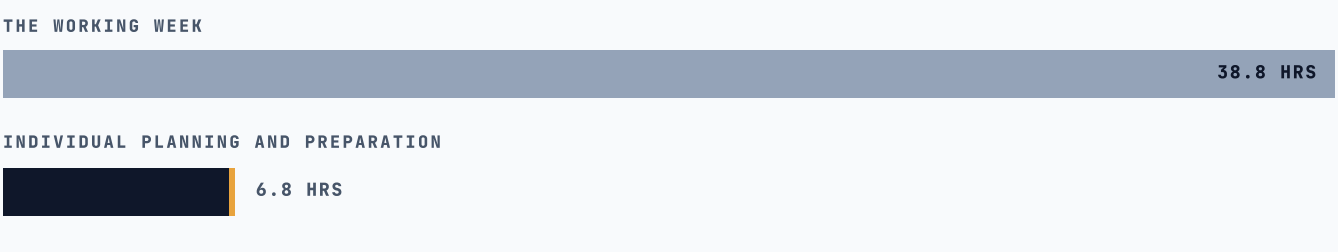
This is why the correct unit of analysis is the lesson. A school can adopt a platform and no lessons change. A teacher can complete a training and no lessons change. But if a single lesson changes, learning changes, because the lesson is where every input in the system, curriculum, teacher capability, materials, time, finally converges on a child.

5 OECD 2019, TALIS 2018 RESULTS, VOLUME I

We propose a simple test for any educational resource, tool, or programme. Call it the tomorrow morning test: could a competent teacher who has never seen this resource before tonight teach it well tomorrow morning, within tonight's preparation budget, with the equipment the school actually has, inside the minutes the timetable actually gives? A resource that passes is classroom-ready. A resource that fails is not a classroom resource yet. It may be a good idea, a promising platform, or an ambition for professional development. It is not something a teacher can use, and the difference is the whole subject of this paper.

The test sounds humble. It is in fact demanding, because it prices in every hidden cost the sector routinely externalises onto teachers: the searching, the adapting, the translating, the printing, the improvising of equipment, the guessing at answers. The evidence suggests those hidden costs are where innovation goes to die.

FIGURE 3 · THE QUARTER HOUR



ONE LESSON'S SHARE · 6.8 HOURS ACROSS ≈26 LESSONS TAUGHT

≈16 MIN

The entire budget a new resource competes for.

SOURCE: OECD TALIS 2018 · DERIVATION SHOWN AT TRUE PROPORTION

03 What the evidence actually rewards

If the last mile framing is right, the interventions that shorten it should outperform the interventions that ignore it. They do, and the margin is not small.

The cleanest demonstration is Jackson and Makarin's randomised trial of off-the-shelf lesson materials.[6] Middle-school mathematics teachers were given complete, ready-to-teach lessons with light supports. Student achievement rose by around 0.08 standard deviations overall, with the largest gains for weaker teachers, at a cost of roughly $430 per teacher. The authors note the approach is more cost-effective than most known interventions aimed at improving teacher quality. The resource did not try to transform the teacher. It removed the last mile, and learning moved.

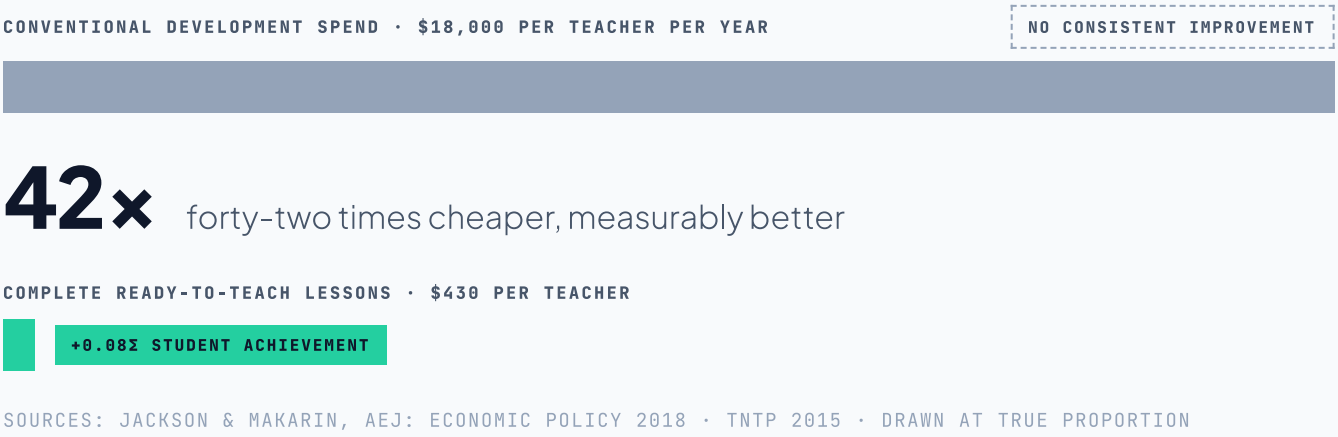
The pattern scales. The Global Education Evidence Advisory Panel, convened by the World Bank, FCDO, and UNICEF, ranks structured pedagogy, meaning complete lesson plans, aligned materials, and supporting training delivered together, among the best buys in global education.[7] Kenya's Tusome programme, built on exactly that architecture of structured teacher guides and matched student materials, moved national early-grade reading outcomes at scale.[8] Reviews of curriculum materials reach the same conclusion from the cost side: switching to high-quality instructional materials rivals far more expensive teacher-effectiveness strategies.[9] And research on effective professional development finds that what works is content-focused, sustained, and anchored in the actual curriculum materials teachers will use.[10] Training that arrives without materials asks teachers to build the bridge themselves.

6 JACKSON & MAKARIN 2018, AEJ: ECONOMIC POLICY 10(3) · 7 GEEAP, SMART BUYS 2020, UPDATED 2023, WORLD BANK, FCDO, UNICEF · 8 PIPER ET AL., TUSOME, RTI INTERNATIONAL · 9 STEINER 2017, STANDARDSWORK · 10 DARLING-HAMMOND, HYLER & GARDNER 2017, LEARNING POLICY INSTITUTE

Teachers have already voted on all of this with the only currency they have. RAND's surveys found that nearly all American teachers assemble instructional materials themselves from search engines and sharing sites,[11] and marketplaces such as Teachers Pay Teachers grew into billion-download ecosystems on the same revealed preference. The demand was never for inspiration. It was for something usable by morning, and where systems failed to supply it, teachers built a grey market.

The field keeps buying transformation. The evidence keeps rewarding preparation.

FIGURE 4 · PREPARATION BEATS TRANSFORMATION



SOURCES: JACKSON & MAKARIN, AEJ: ECONOMIC POLICY 2018 · TNTP 2015 · DRAWN AT TRUE PROPORTION

11 OPFER, KAUFMAN & THOMPSON 2016, RAND CORPORATION

04 The longest last mile: teaching in Arabic

Every gap described so far widens when the teaching language is Arabic.

Start with supply. Arabic is a first language for more than 400 million people, yet it remains dramatically under-represented in open educational resources, and much of what exists is translated rather than authored. Translation moves words across the last mile and leaves everything else behind: the examples still describe other people's rivers, the data still measures other people's cities, the safety notes still assume other people's equipment. A climate lesson built on Alpine glaciers does not fail a student in Aseer because the science is wrong. It fails because nothing in it belongs to her, and belonging is what a 40-minute lesson runs on.

Now add commitment, because the region is not short of it. Saudi Arabia's 2026 Voluntary National Review commits, on page 154, to sustainability training for 250,000 teachers by 2028, and names the constraint itself: varying teacher readiness in delivering sustainability and environmental education.[12] Our audit of the national professional development catalogue found 77 accredited courses, none covering sustainability.[13] National teacher platforms carry hundreds of thousands of users a day. The ambition is published, the channel exists, and the classroom-ready resource layer between them is missing. This is the last mile drawn in national scale: a signed commitment at one end, a Sunday-morning lesson at the other, and the shortest, hardest distance in education in between.

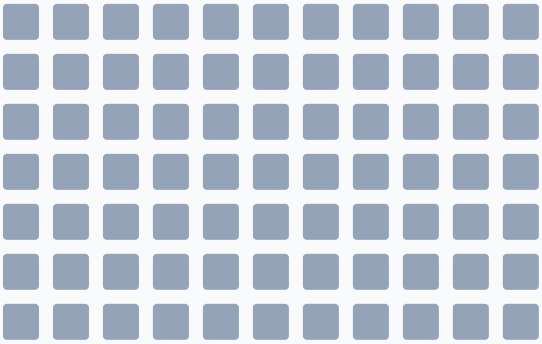
12 KINGDOM OF SAUDI ARABIA 2026, VOLUNTARY NATIONAL REVIEW, P.154 · **13** GBOX 2026, NATIONAL CATALOGUE AUDIT

FIGURE 5 · A SIGNED COMMITMENT, AN EMPTY SHELF

250,000

teachers to be trained in sustainability by 2028, committed in the Kingdom's Voluntary National Review, p.154

77 ACCREDITED COURSES IN THE NATIONAL PROFESSIONAL DEVELOPMENT CATALOGUE



SUSTAINABILITY COVERAGE: 0

SOURCES: KSA VNR 2026, P.154 · GBOX CATALOGUE AUDIT 2026 · ONE MARK PER COURSE

We treat the Saudi case throughout this series not as a deficit story but as the opposite: the place where the last mile will be walked first, deliberately, at scale, in Arabic, and measured. The remaining papers take up the regional readiness gap, climate learning designed for arid realities, and the evidence framework the walk requires.

05 The Classroom-Readiness Standard

If the last mile is the problem, the sector needs a way to see it before purchase, not after failure. What follows is a scoring standard with three design rules. It measures the resource, not the teacher, because the resource is what a buyer controls. It is scorable from the artefact itself, by a non-specialist, in under an hour, because a standard nobody can apply is a press release. And it prices hidden costs deliberately, because hidden costs are where the last mile hides.

Twelve criteria, four domains. Score each criterion 0 (absent), 1 (partial), or 2 (fully met), for a maximum of 24.

DOMAIN A · FIT THE HOUR

01 **Whole-lesson coverage.** The resource fills the national lesson length end to end, as a complete teachable sequence, not a list of activity ideas.

02 **Preparation inside the budget.** Total teacher preparation, printing included, fits within 30 minutes.

03 **Closed materials list.** Every physical item is named and counted, and is either included with the resource or verifiably standard in the target school.

DOMAIN B · FIT THE TEACHER

04 **Zero-training deployment.** A competent teacher can deliver it without prior professional development; all guidance lives inside the resource itself.

05 **Script where stakes are high.** Verbatim support is provided where it matters most, key questions, transitions, safety steps, with professional freedom everywhere else.

06 **Answers and misconceptions included.** Expected student responses and the common errors are anticipated in the teacher guide, so the teacher is never guessing in front of the class.

DOMAIN C · FIT THE CONTEXT

07 **Curriculum-coded.** Mapped to the official code of the specific unit it delivers, not loosely aligned with a subject.

08 **First-language first.** Authored in the language of instruction, with any second language as a mirror, never the reverse.

09 **Locally evidenced.** Examples and datasets come from the students' own region, and every source is named.

DOMAIN D · FIT THE EVIDENCE

10 **Visible learning output.** The lesson itself produces an artefact or performance that demonstrates learning, inside the period.

11 **Built-in check.** A check for understanding of five minutes or less, with marking guidance, is part of the resource.

12 **Feedback loop.** A working channel exists for teacher usage and outcome data to return to the maker.

Readiness bands. 20 to 24: classroom-ready. 14 to 19: adaptable, meaning usable only with teacher investment the maker has externalised. Below 14: not yet a classroom resource, whatever else it may be.

The bands are deliberately blunt. A 13 is not a weak product. It is a product that has quietly hired the teacher as its unpaid finishing department.

The Classroom-Readiness Standard

Score each criterion 0, 1 or 2 from the resource itself. Maximum 24.

A · FIT THE HOUR

01
Whole-lesson coverage
Fills the national lesson length end to end

02
Preparation inside the budget
Preparation, printing included, within 30 minutes

03
Closed materials list
Every item named, counted, included or standard

B · FIT THE TEACHER

04
Zero-training deployment
All guidance lives inside the resource itself

05
Script where stakes are high
Verbatim support for questions, transitions, safety

06
Answers and misconceptions included
Expected responses and common errors anticipated

C · FIT THE CONTEXT

07
Curriculum-coded
Mapped to the official code of the unit it delivers

08
First-language first
Authored in the language of instruction

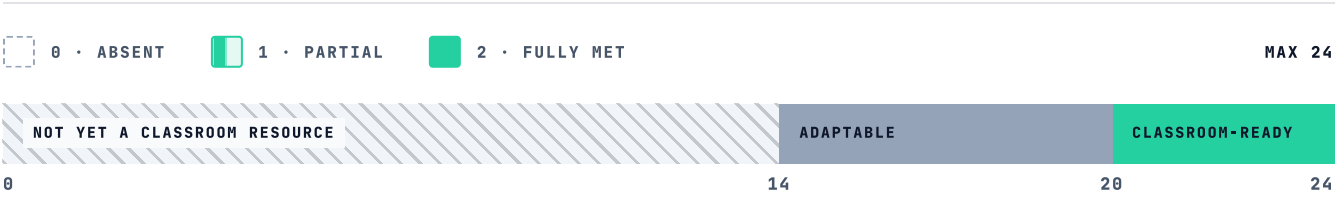
09
Locally evidenced
Examples and datasets from the students' own region

D · FIT THE EVIDENCE

10
Visible learning output
An artefact or performance inside the period

11
Built-in check
A five-minute check with marking guidance

12
Feedback loop
Usage and outcome data return to the maker

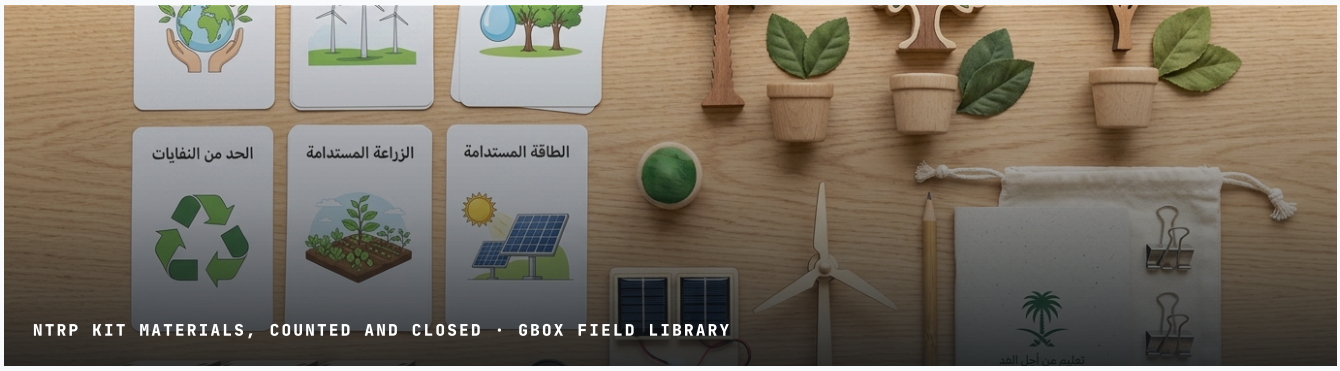


06 The standard in use

A standard earns trust only when its author scores first. We applied the CRS-12 to the flagship kit in our own National Teacher Readiness Programme library, Water from Air, which turns measured fog-harvesting data from the Aseer highlands into two 40-minute science lessons for Grade 4, Ministry code SCI-4-2.1.

Walking the criteria: the kit ships as a complete two-lesson sequence with an integrated teacher guide, slides, source data and documentation, a quick comprehension check, and two short films covering experiment set-up and the classroom story, authored Arabic-first with an English mirror. Its physical field box equips a class of 30 across five student workstations, every item counted, from double-mesh collector frames to graduated cylinders and 30 laminated record cards. The lesson produces a measured, recorded output, millilitres of water per spray cycle, inside the period. On our internal scoring the kit reaches the classroom-ready band, with the honest gap sitting in Domain D: the feedback loop exists in the beta library but does not yet return outcome data at the depth criterion 12 demands.

Publishing our own gaps is the point. A readiness standard that only ever awards its author full marks is marketing. One that shows where its author falls short is an instrument, and instruments are what the sector is missing. The full criterion-level audit of all 44 kits in the bank publishes alongside this paper.



NTRP KIT MATERIALS, COUNTED AND CLOSED · GBOX FIELD LIBRARY

We invite three groups to score with us. Makers: score your own product and publish the number on the box. Funders: request a CRS score with every proposal, exactly as you request a budget. Ministries and registries: record the band beside every listed resource, so a teacher browsing at 9pm can see in one glance what tomorrow morning will cost her.

FIGURE 7 · ONE KIT, SCORED IN PUBLIC

WATER FROM AIR · ASEER · GRADE 4 · SCI-4-2.1

Criterion 12, feedback loop: beta library live; outcome depth not yet met

INTERNAL DRAFT · SCORES TO CONFIRM BEFORE PUBLICATION

SOURCE: GBOX NTRP KIT AUDIT, 2026

07 What changes if the buyer changes

Standards move markets when buyers adopt them, so the closing argument is addressed to the people who sign.

For funders and CSR leaders, the move is to make the readiness band a condition of funding, and preparation time a reported metric, so that every riyal and dollar is buying lessons taught rather than licences issued. The difference between funding a 21 and funding a 12 is not quality in the abstract. It is whether the grant produces classroom hours or shelf inventory.

For ministries and platform owners, the move is procedural: score before listing, publish the band, and let the registry itself teach the market what classroom-ready means. A national platform that displays readiness scores does more for resource quality in a year than a decade of exhortation, because makers build to whatever the shelf rewards.

For makers, including us, the move is to treat the standard as the design brief from day one. Write the lesson before the platform. Count the materials before the launch. Put the answers in the guide, the code on the cover, and the local data in the dataset, and accept that the 30-minute preparation budget is not a constraint on the product. It is the product.

Education innovation does not need another decade of promising pilots. It needs resources that survive contact with tomorrow morning, a way to recognise them before purchase, and buyers willing to insist on the difference. The last mile is short. It is time the sector walked it.

THE THREE MOVES, ONE LINE EACH

FUNDERS

make the readiness band a condition of funding

MINISTRIES

score before listing, publish the band on the registry

MAKERS

treat the standard as the design brief from day one

FIGURE 8 · WHAT THE SHELF REWARDS

A reinforcing loop: every published band teaches the market, and the market builds the next resource closer to classroom-ready. The system needs no enforcement, only a shelf that shows the number.

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. GBOX's National Teacher Readiness Programme (NTRP) and Teacher Resource Library put the argument into practice in Saudi Arabia.

GREENBOX.WORLD/NTRP

Sources

01 TNTP (2015). The Mirage: Confronting the Hard Truth About Our Quest for Teacher Development.

02 Cristia, J., Ibarrarán, P., Cueto, S., Santiago, A., Severín, E. (2012). Technology and Child Development: Evidence from the One Laptop per Child Program. Inter-American Development Bank.

03 Glimpse K12 (2019). Analysis of K-12 school spending and software licence utilisation.

04 LearnPlatform by Instructure (2022). EdTech Top 40 and district tool-usage reporting.

05 Jackson, C. K., Makarin, A. (2018). Can Online Off-the-Shelf Lessons Improve Student Outcomes? Evidence from a Field Experiment. American Economic Journal: Economic Policy, 10(3).

06 Global Education Evidence Advisory Panel (2020; updated 2023). Smart Buys: Cost-Effective Approaches to Improve Global Learning. World Bank, FCDO, UNICEF.

07 Piper, B., et al. Evaluations of the Tusome Early Grade Reading Activity, Kenya. RTI International.

08 Steiner, D. (2017). Curriculum Research: What We Know and Where We Need to Go. StandardsWork.

09 Darling-Hammond, L., Hyler, M. E., Gardner, M. (2017). Effective Teacher Professional Development. Learning Policy Institute.

10 OECD (2019). TALIS 2018 Results, Volume I: Teachers and School Leaders as Lifelong Learners.

11 Opfer, V. D., Kaufman, J. H., Thompson, L. E. (2016). Implementation of K-12 State Standards for Mathematics and English Language Arts and Literacy. RAND Corporation.

12 UNESCO (2024). Global Report on Teachers: Addressing Teacher Shortages.

13 Kingdom of Saudi Arabia (2026). Voluntary National Review, p.154.

14 GBOX (2026). Audit of the national professional development catalogue: 77 accredited courses, sustainability coverage.

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.

Delivery is the product.

The last mile is short. It is time the sector walked it.

NEXT IN THE SERIES · 02 · THE REGIONAL READINESS GAP