



Scaling Teacher Performance Contracts

An Adaptive Experimental Study in Rwanda



Clare Leaver, Arnesh Chowdhury , Leodomir Mfura , Owen Ozier , Pieter Serneels , Andrew Zeitlin

Proof of Concept Isn't Proof at Scale

A prior RCT showed pay-for-performance teacher contracts roughly doubled annual learning gains. But moving delivery from researchers to routine government systems risks a “**voltage drop**” — effects that attenuate outside controlled conditions.

- **High-dimensional design space**

Which contract features actually matter, and which are safe to drop?

- **Unknown interactions**

Do features complement or substitute for one another?

- **Few shots at adaptation**

Political & logistical constraints limit how often you can experiment.

- **Slow-materializing outcomes**

Learning effects can take years to show up in the data.

“Policy optimization in a rugged landscape”

Pritchett, Samji, & Hammer (2013) suggest a "rugged" fitness landscape limits learning.

We posit that a Bayesian approach to analysis, built on top of a hub-and-spoke exploration strategy that starts from the tested model, can help identify impactful demands when the landscape is not perfectly rugged -- that is, when similar designs tend to perform similarly

Built on a Proven Foundation



The original STARS trial (2015–2018)

IPA & gui2de tested pay-for-performance contracts in 164 schools across 6 districts — bonuses up to 5% of salary, split 50/50 between student learning gains and teacher inputs (presence, preparation, pedagogy).

0.20 SD

combined learning effect
by Year 2 (effort + composition)

≈ 1 year's

learning gain — the effect size
matched a full extra year

Why now: The policy window

Imihigo system

- Rwanda's civil-service performance contracts already allow up to 5% variable pay — but teacher scores were near-universal.
- Government keen to build upon the success of the STARS RCT

2023 salary reform

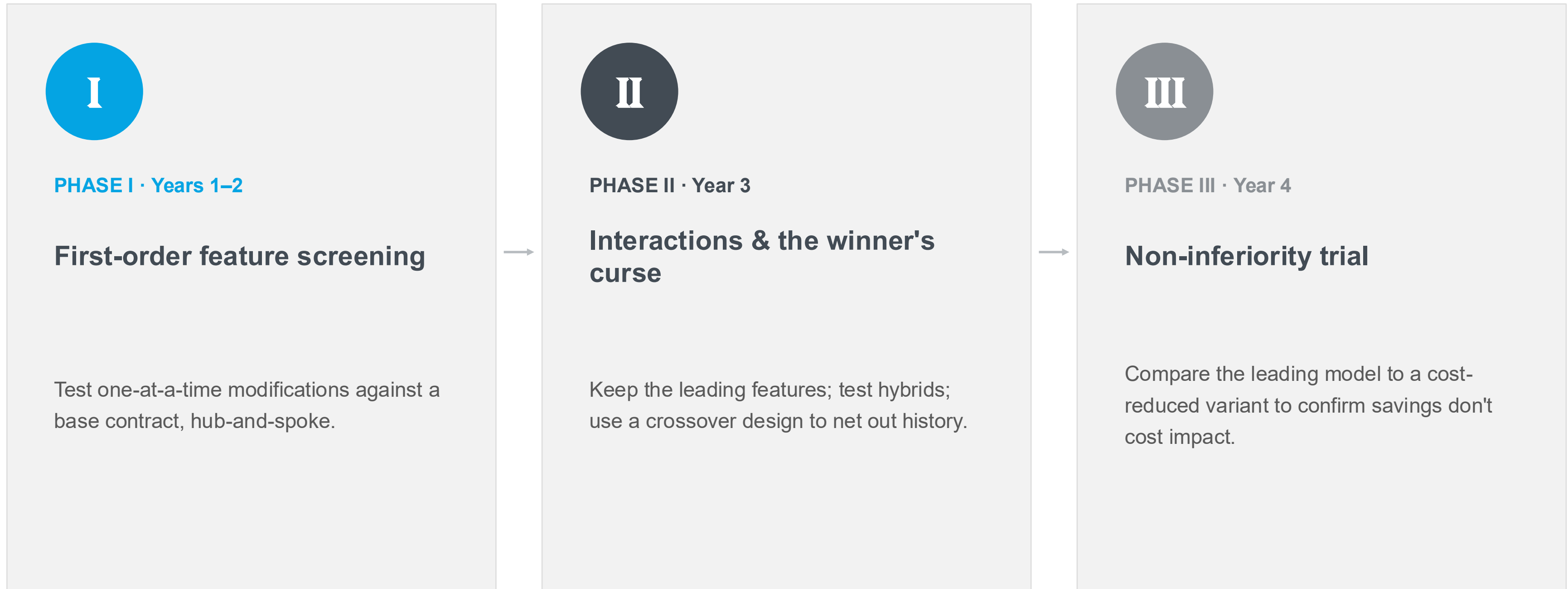
- Teacher salaries nearly doubled, creating appetite to link pay meaningfully to performance.

CAMIS rollout (2022)

- First nationwide teacher-level learning database in Sub-Saharan Africa — made linking pay to learning feasible.

A Phased, Adaptive Approach

Each phase answers a different question, narrowing the design space as evidence accumulates.



Screening features against a base Contract



Base Model A: the “5Ps”

- 1 Presence (unannounced SEI visits)
- 2 Preparation (lesson plans)
- 3 Pedagogy (classroom observation)
- 4 Participation in CAMIS
- 5 Pupil learning gains (pay-for-percentile)

Composite score → “stoplight” payout: 5% bonus above 80, 3% between 70–80, 0% below 70.

Four one-at-a-time variations (Cohort 1)

B1	Head Teacher subjective assessments of the 5Ps
B2	Termly feedback on relative performance rankings
B3	Alternative threshold payout scheme
B4	Re-testing 5% of students to audit learning

248 treatment + 95 control schools (Cohort 1); +150 schools added in Year 2 (Cohort 2), introducing Model B5.

Strong input effects; two arms lead on learning



Student learning: two designs emerge as leaders (Year 2)



No evidence of gaming of administrative assessments : High-stakes CAMIS scores track independent low-stakes IPA assessments closely, in both treatment and control schools.

Phase2: Testing Complementarities with a Crossover

Models A/B3 and B2 were discontinued. Cohort 1 schools from those arms were re-randomized into the surviving and hybrid arms; Cohort 2 schools continued under Model A or B5.

- Provides a proper experimental counterfactual for the new hybrid models
- Lets researchers separate contemporaneous effects from prior treatment history — important given dynamic effects seen in Phase I

New Head Teacher incentive: Model B5

Head Teachers get their own imihigo contract — 50% of their score comes from the school average of teachers' 5P scores.

Hybrid models

Model B6

B1 + B5

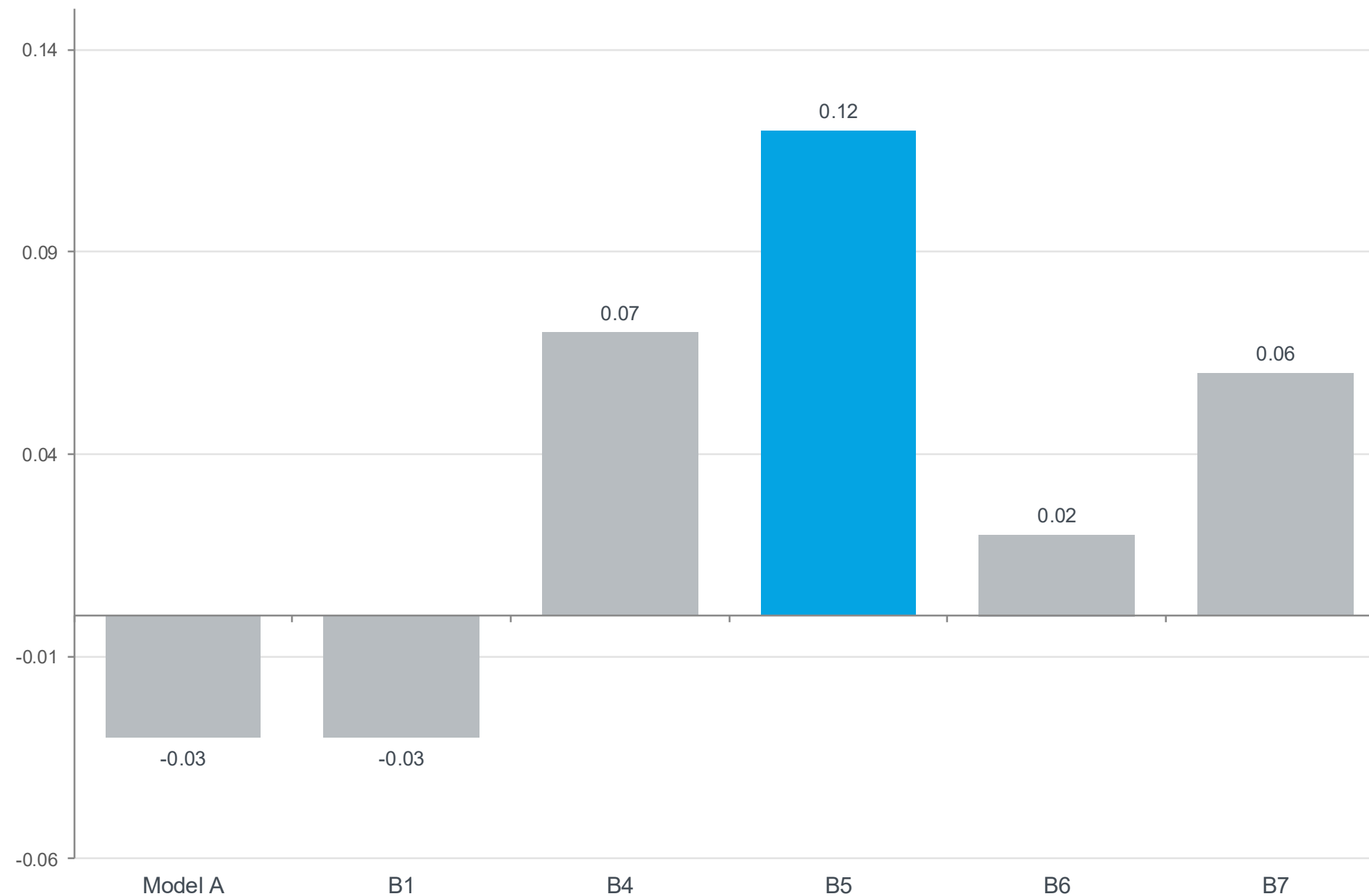
Head Teacher assessment of teachers, combined with the Head Teacher's own imihigo incentive.

Model B7

B1 + B4 + B5

HT assessment + student re-testing audits + the Head Teacher's own imihigo incentive, all combined.

Phase 2: Model B5 Delivers the Largest Learning Gain



Pooled student-learning impact by treatment arm (independent, low-stakes assessments)

MODEL B5 · HEAD TEACHER INCENTIVES

0.12 SD

pooled learning gain ($p = 0.03$) $\approx$ 0.6 additional school years

- English: +0.15 SD, Math: +0.14 SD
- Similar gains in lower & upper primary
- Largest pedagogy improvement too (+0.26 SD)

HT monitoring alone (B1) fades by Year 3; layering it onto B5 ($\rightarrow$ B6) doesn't help. The incentive channel, not the monitoring channel, drives results.

Impacts Achieved with Low Per-Pupil Implementation Costs and Salary Savings



Three families of cost-effectiveness questions and estimands

Estimand	Notes and assumptions	Results
Ex-post CEA during A/B testing period	● Well identified benefits; per-arm CEA requires assumptions about attribution of fixed costs ● Aggregates effects of all arms so not aligned with exploration motive, and not a well-defined intervention	Benefit-cost ratio 14.3 Marginal Value of Public Funds: 23.4 Social Rate of Return 26.9%
At-scale operational CEA	● Simplifying assumption to count only benefits for (and attribute all costs to) a single cohort in each year	BCR 1979 MVPF infinite SROR infinite LAYS per \$100: 214
Return on funder investment	● Requires assumption about duration of implementation post scale-up (we use 5 years). ● Includes all research costs	BCR 67.9 MVPF 203.8 SROR 35.9%

Notes: 1. Assume 5% discount rate, 17.75% earnings gain per SD. 2. Family 2: Direct costs per pupil are \$0.64 but *imihigo* savings per pupil are \$0.449 due to reduced grade inflation.

From Trial evidence to National Policy



2025/2026: The Government of Rwanda began a nationwide rollout of learning-based imihigo for both teachers and Head Teachers — built on Model B5.

Lessons for adaptive experimentation in policy settings

Iterative testing/Features screening

Rules out weak features before committing power to testing interactions.

Crossover design

Separates true treatment effects from exposure-duration confounds.

Government-embedded

Close Task Force partnership kept design both rigorous and politically feasible.

Teacher performance contracts, adapted through evidence and delivered through government systems, can produce meaningful, cost-effective learning gains at national scale.