


12 randomized A/B tests optimizing tutoring for scale

 Noam Angrist, Colin Crossley, Claire Cullen

 The link to the academic paper [is available here](#).

Tutoring is one of the most effective educational interventions globally, but its high cost has limited scalability. To address this challenge, we used A/B testing—rapid, randomized experiments comparing modified versions of a program—in order to iteratively optimize a phone call tutoring model for greater cost-effectiveness. Seven of 12 tests generated large efficiency gains. Results suggest that the social sector can successfully utilize A/B tests to address both sides of the scaling equation, reducing costs and increasing effectiveness.

Despite growing school enrollments around the world, millions of children still have not acquired foundational skills in literacy and numeracy (World Bank 2018; Angrist et al., 2021). One of the most effective educational approaches is tutoring, yet high costs have remained a barrier to scale (Kraft et al., 2022).

To address this challenge, Youth Impact conducted **12 A/B tests** to further optimize a tutoring program delivered through phone calls. We optimized a proven approach, shown to improve learning in prior RCTs conducted in six countries, for greater cost-effectiveness and scalability. A/B testing has become a common approach in the technology sector (Kohavi et al., 2020), but has yet to see wide use for social programming, despite its potential to help address key scaling constraints.

Iterative A/B testing compares two randomized groups—groups A and B—which are equal except for one difference: group A is the status quo program and group B is a slightly modified version of the program. A/B tests can act as a bridge between RCTs, which often ask the question “does the program work?”—and ongoing implementation questions, which ask “what works even better and cheaper?”

A/B testing is characterized by three Rs: Rigorous, Rapid, and Regular, described in Box 1. These randomized, fast-cycle experiments allow for real-time, cumulative, and continuous learning.

Box 1

What are the three Rs?

Rigorous



A/B tests follow an experimental design, randomizing students into groups to detect causal impacts.

Rapid



Testing can occur quickly and cheaply. Ideally, A/B tests can be delivered termly.

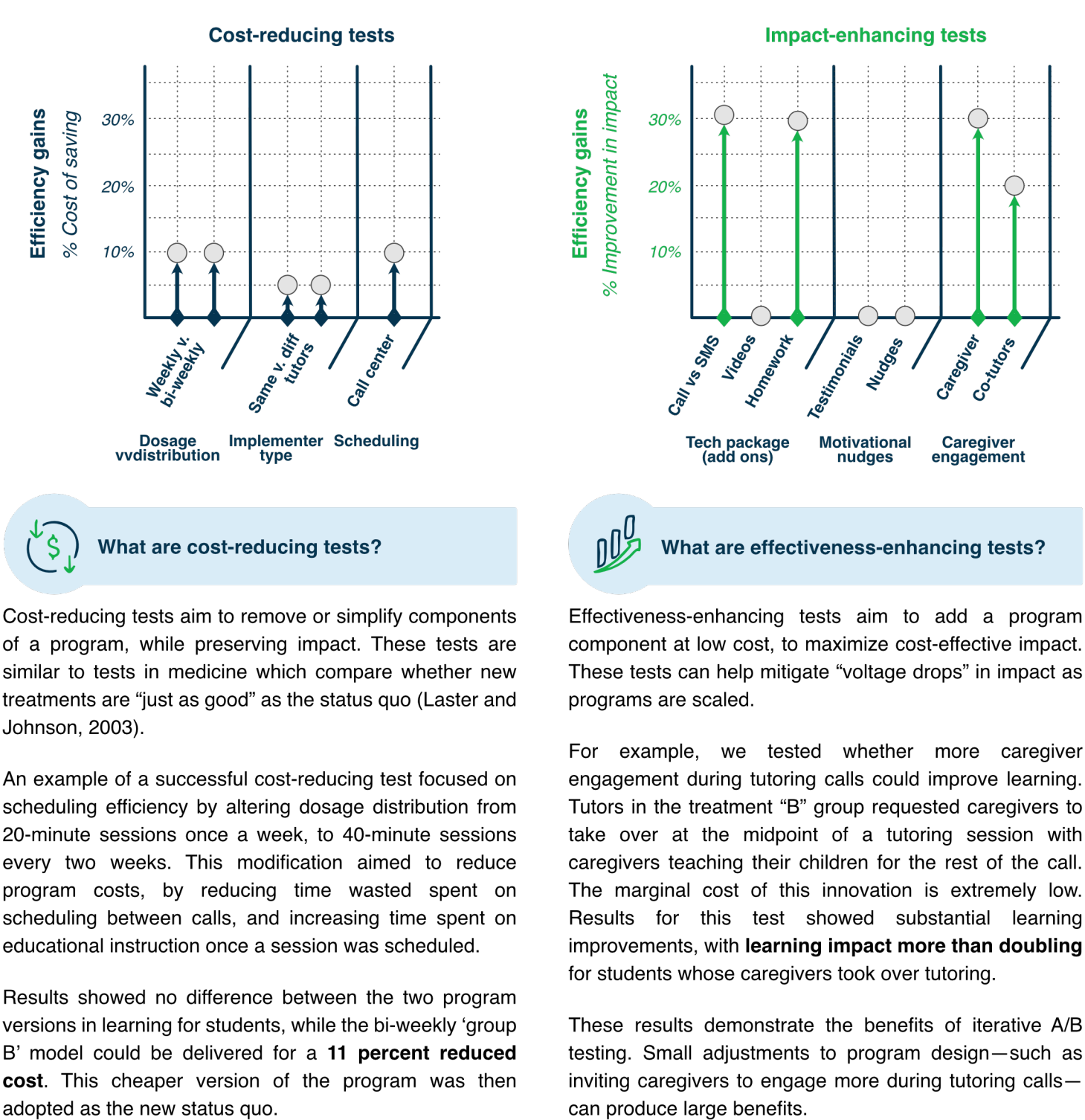
Regular



Testing is an ongoing part of a cumulative learning agenda, embedded in organizational structures and systems.

Results showed large efficiency gains. **Seven of 12 A/B tests improved cost-effectiveness**, with efficiency gains up to 30 percent per test. These tests are summarized in Figure 1 and can be grouped into two categories: **cost-reducing tests** as well as **effectiveness-enhancing tests**. Cost-reducing tests assess if a lower cost program can be simplified and streamlined, while remaining as effective as the status quo program. Effectiveness-enhancing tests examine margins to enhance impact at low marginal cost.

Figure 1: Twelve A/B tests showed a range of efficiency gains up to 30 percent per test.



Box 2

Dosage distribution

Group A: 20-minute tutorials every week

Group B: 40-minute tutorials every two weeks

Box 3

Caregiver engagement

Group A: Status quo facilitator-led tutorials

Group B: Encouragement to caregivers to co-lead tutoring



A/B tests are a powerful tool for innovation in the social sector

The 12 A/B tests summarized in this brief demonstrate that simple modifications with low marginal costs can substantially improve the impact of a program, especially as efficiency gains accumulate across multiple tests over time.

First, seven of 12 tests led to measurable efficiency gains—a “hit rate” that exceeds the tech sector benchmark of between 10 and 40 percent. This demonstrates that **iterative testing in social programs can yield high returns**.

Second, successful tests achieved **up to 30 perfect efficiency improvements per test**, including both cost reductions and effectiveness enhancements. This optimization can help mitigate and even reverse the typical “voltage drop” seen when scaling social programs (List 2022).



Some tests, such as caregiver engagement, lead to extremely high cost-effectiveness

One of our most impactful A/B tests involved encouraging caregivers to co-lead tutoring calls. This program innovation came at an additional cost of no more than \$0.48 per child, and **more than doubled learning outcomes**. At a marginal cost of under fifty cents per child, this adjustment yielded a learning gain so large that, if evaluated independently, it would rank among the most cost-effective interventions in the education literature.

Measuring practitioner prior and posterior belief showed that A/B testing corrected implementers’ misperceptions, making decision-making more evidence aligned. This reinforces A/B testing as a tool for **improving how organizations learn on the frontlines of implementation**.



Practitioners update beliefs on what works based on A/B testing results

A growing number of organizations are starting to implement A/B testing regularly. Many international development groups, including the Foreign, Commonwealth & Development Office (FCDO), the Global Education Evidence Advisory Panel, and the What Works

Hub for Global Education, promote evidence-based decision making. Adding A/B testing to the evaluation toolkit of researchers, policymakers, and implementers offers a promising way forward.

Implementing & research coalition



What Works Hub
for Global Education



Funding partners



UBS Optimus Foundation

Contact ccullen@youth-impact.org or visit youth-impact.org for more information about A/B testing.