March 2026

Innovations for Poverty Action

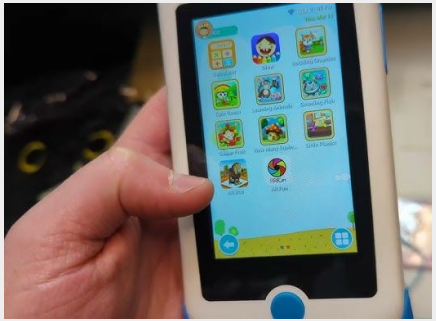
A Practical Approach to Developing A/B Testing Systems for Digital-first Organizations

Andrés Parrado

Most interventions suffer a “voltage drop” at scale

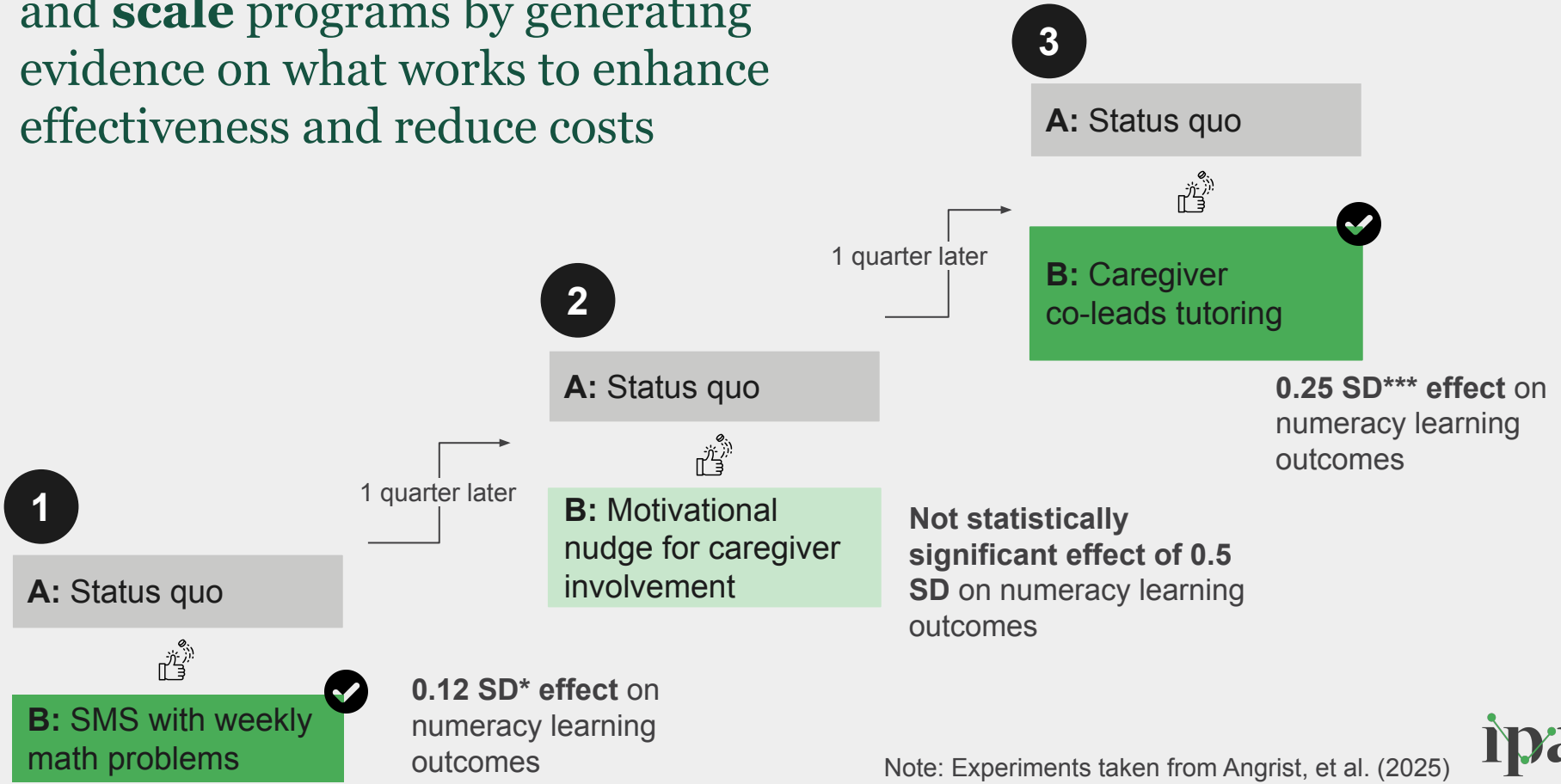
Between 50-90% of programs experience a “voltage drop” when scaled, where the benefit-cost ratio shrinks when moving from a small-scale trial to larger, real-world implementation (List, A., 2023)

Costs increase at scale



Effectiveness decreases with scale

Iterative A/B testing can help **optimize** and **scale** programs by generating evidence on what works to enhance effectiveness and reduce costs



Note: Experiments taken from Angrist, et al. (2025)

Not every test will be a hit

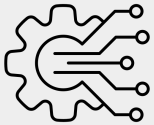
??

Maybe as few as 10% (cited in Angrist, et al. (2025))

A/B testing can help address scaling challenges

But, how to do it right?

Invest in tests that have a **high potential return** to enhance program cost-effectiveness



Develop the **capabilities and systems** for **iterative** A/B tests

ipa

How to design & prioritize high potential A/B tests?

1. Identify what do you want to enhance: **reach, efficacy, or cost**

2. Identify the **metric** you will use to test

3. Identify and prioritize **evidence-based variations** to test out

1 Identifying what to optimize: Look for opportunities to improve interventions' cost-effectiveness

$$\frac{\text{Program's reach} \times \text{Program's impact}}{\text{Program's costs}} = \text{Program's Cost-effectiveness}$$

2 Identify key metrics for A/B testing

Cost-effectiveness lever	Example metrics
Reach	● # active users ● # onboarded users
Impact	Product KPIs ● **User engagement:** Weekly frequency of usage
	Theory of Change outcomes ● **Knowledge:** Changes in foundational numeracy skills ● **Attitudes/perceptions:** Changes in teacher's perception of the importance of addressing learning variability in the classroom ● **Behaviors:** Changes in teacher practices
Cost	● Unit cost of program delivery

3 Identify and prioritize variations to test

Collect data to inform what to test

- Literature review
- User research
- Interviews with experts/field staff

Define variation options

- 1: Send nudging reminders
- 2: Call students daily
- 3: Create a “streak” feature
- 4: Written commitment

Prioritize variations and sequence them

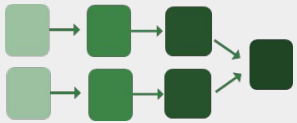
- ***Relevance:*** How likely is this variation to improve the key outcome metric?
- ***Implementation feasibility:*** How easy and/or costly will it be to implement this variation?

At the end you should have a road map of possible A/B tests

Illustrative example: Designing a set of A/B tests for Shaia (teacher support chatbot)

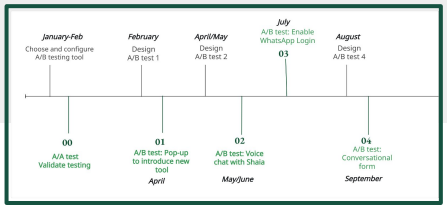
Mentu decides they want to test things that could enhance the:

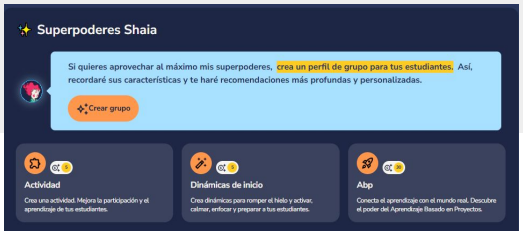
Program's impact



Mentu identified **the use of the learning variability feature** as the key metric.

Articulated a **series of A/B tests** into a roadmap





What capabilities and systems are needed to run an A/B test?

1. Define sample size

2. Set-up the necessary data systems & tools

3. Analyze results and make decisions

Key enablers for organizations to adopt and implement A/B testing

Sample size considerations

- **A stable and sufficiently large user base**, to support rigorous comparisons.

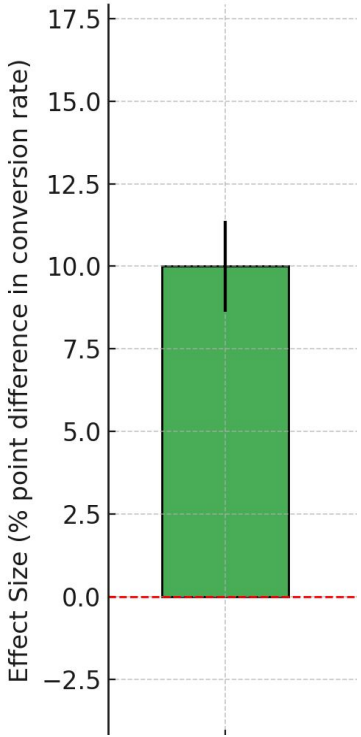
Systems and tools

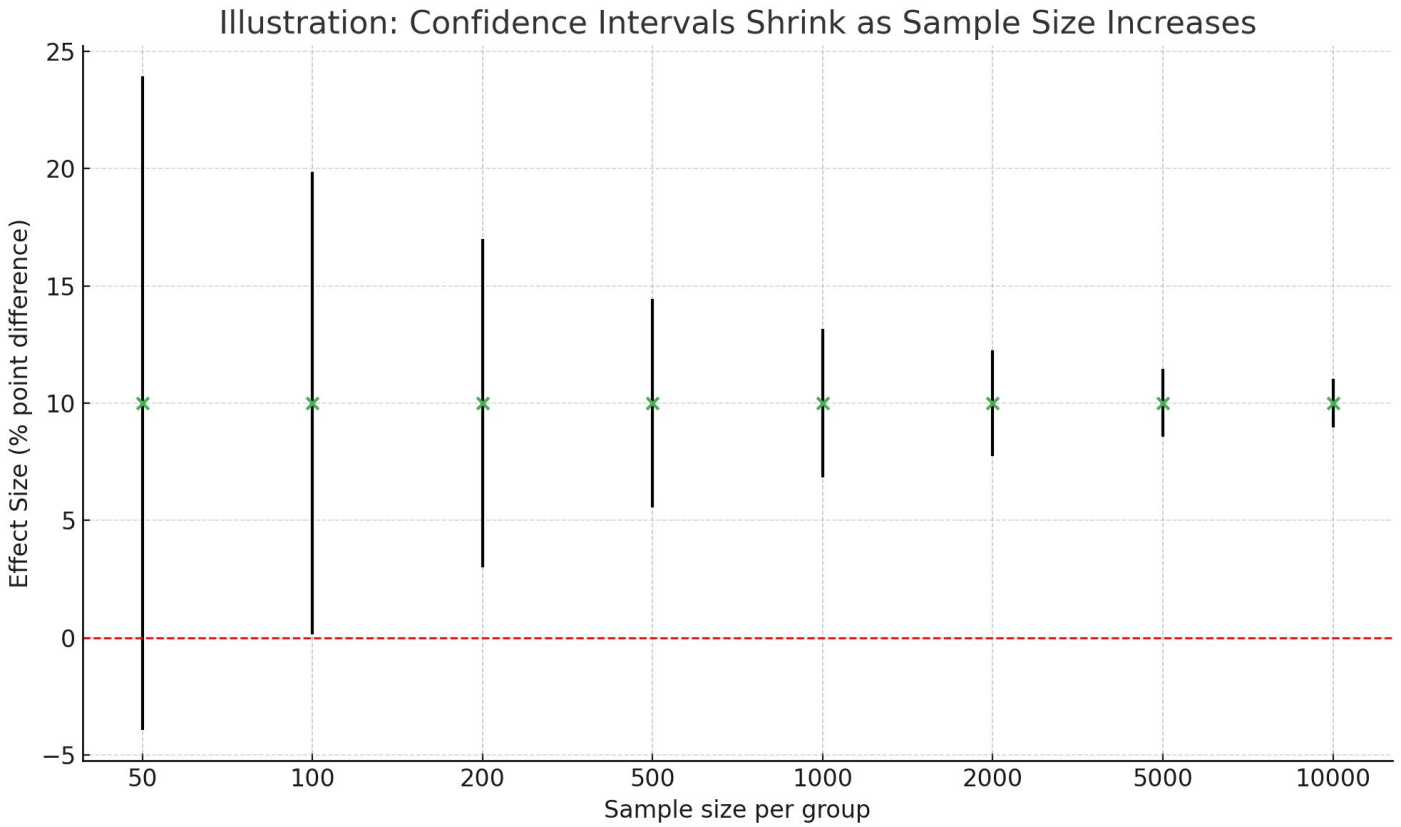
- **Reliable data systems** to track relevant outcomes.
- **Experimentation tools**, that allow for random assignment and multiple test management.

Analytical capacity

- **Strong data management capacity,** to connect data infrastructure with experimentation tools.
- **Statistical and analytical capacity** to design valid experiments and accurately interpret results.

1 Key concept: Larger sample size reduce uncertainty of results





2 What systems & tools are needed for A/B testing?

The tool selection depends on the nature of the program/ product

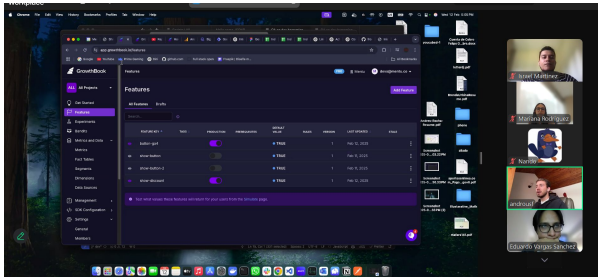
Experimentation tool

- Experiment setup: segment target sample and randomly assign users
- Deploy variations

Data Infrastructure should allow for the integration between these systems

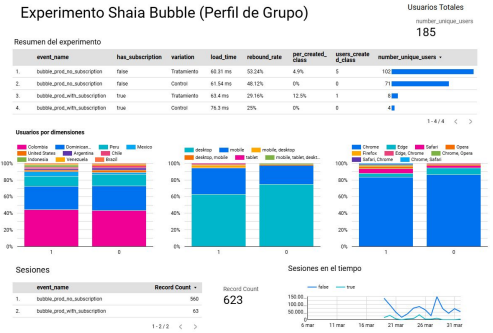
Data Analytics tool

- Capture and visualize key metrics
- Statistically analyze results



GrowthBook (open source)

Other tools specifically for the social sector are **Evidential** and **UpGrade** (education)



Looker Studio

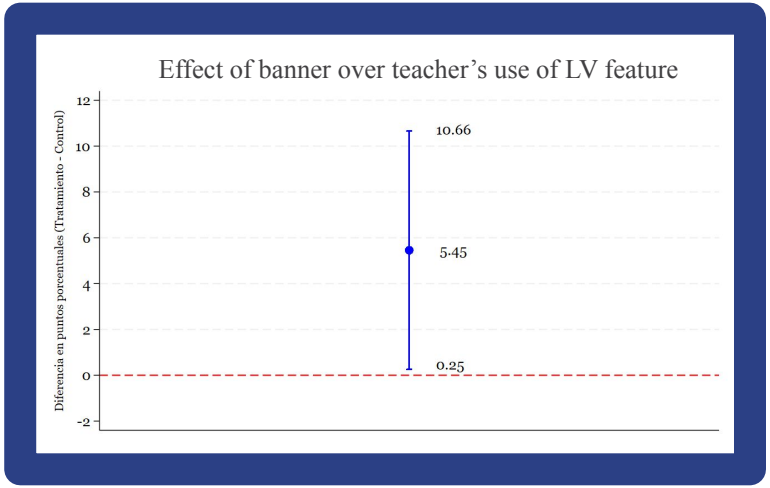
3 Analyzing A/B tests results and making decisions

Assess the **credibility** of the experiment:

- **Sample size** necessary to detect a meaningful effect
- **Groups are comparable** on key characteristics

Assess the **effectiveness** of the variation:

- Is the difference between groups **statistically significant?**
- Is the difference large enough to matter?



Large effect with non statistically significant results?
Assess the level of uncertainty (CI) in the results

Interpret the results and **make a decision**

Inconclusive or null: repeat the experiment or discard variation

Positive effect: scale the variation to all users or continue testing to find a more effective variation

Negative effect: discard the variation and use findings to inform future iterations

Happy testing!

Innovations for Poverty Action

Thank you!

Contact us:
Andrés Parrado, Associate Director
Right-fit Evidence Unit
aparrado@poverty-action.org

poverty-action.org