EdTech Hub

Clear evidence, better decisions, more learning.

Experiment types for EdTech

UKaid
from the British people

THE WORLD BANK

unicef
for every child

About this document

Recommended citation	Rahman, A., Carter, A., WambûiKuria, C., Mbugua, C., Simpson, L., & Kably, N. (2022). *Experiment Types for EdTech*. Working Paper. https://doi.org/10.53832/edtechhub.0128 Available at https://docs.edtechhub.org/lib/X8MUXS39. Available under Creative Commons Attribution 4.0 International, https://creativecommons.org/licenses/by/4.0/.
Licence	Creative Commons Attribution 4.0 International https://creativecommons.org/licenses/by/4.0/. You — dear readers — are free to share (copy and redistribute the material in any medium or format) and adapt (remix, transform, and build upon the material) for any purpose, even commercially. You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
Notes	EdTech Hub is supported by UK aid (Foreign, Commonwealth and Development Office), Bill & Melinda Gates Foundation, World Bank, and UNICEF. The views expressed in this document do not necessarily reflect the views of UK aid (Foreign, Commonwealth and Development Office), Bill & Melinda Gates Foundation, World Bank, and UNICEF.

Experiment Types

Why this tool?

An EdTech Hub sandbox fast-tracks promising EdTech interventions by providing funding, tools, and access to evidence.

It provides a space for partners to test and grow ideas in conditions of uncertainty.

Since 2020, EdTech Hub has worked with partners in eight countries to test and grow EdTech based on our sandbox methodology.

When testing and growing EdTech, this tool provides a non-exhaustive list of ways to design and run experiments.

EdTech Hub

Why do we experiment?

When we don't know whether something is 'true' or not we experiment.

Experiments give us stronger (but not 100% accurate) knowledge on whether our ideas will work. And they allow us to learn about where and what we need to change in order to improve.

Experiments also:

- Save time and money needed in order to learn.
- Generate measurable and clear results.

"Make sure
you are building
the **right it** before
you build **it right**"

Alberto Savoia

🤓 Read more: *Pretotype It* by Alberto Saviola

What makes an experiment a good experiment?

> "Plan your tests and **test your plans.**"
>
> **Peter Murray & Steve Ma**
> The Promise of Lean Experimentation, Stanford Social Innovation Review

Bad experiment

"We'll talk to people and see whether they like the idea."
"We'll spend 6 months in strategic planning mode."
"We'll spend 6 months in design or build mode."
"We'll know it's worked when we see it."

Good experiment

Be specific: "We'll create a mock-up and give it to X people. We'll measure how many actually use it, which gives us Y. Then we'll measure how many still use it after four weeks, which is Z. The gradient from X to Y to Z will tell us about expressed interest, actual interest, and sustained interest."

Be lean: "We don't need to do that yet. It's not the most critical thing. Let's have just enough functionality to test what's critical."

Experiment types

1. **The Automaton Chess Player**

2. **The Pinocchio**

3. **The Minimum Viable Product**

4. **The Provincial**

5. **The Fake Door**

6. **Time, Money, or Reputation**

7. **The Pretend-To-Own**

8. **The Lean Cost Model**

9. **The Re-Label**

10. **The Mini-RCT**

11. **The Multivariate Test**

1. The Automaton Chess Player

Before investing or building complex tech, simulate it with real people and see if it does what you think.

_____

+ Allows simulation of a product/service but without sinking your budget into building the full 'back end'.

+ Generates real-use data — to the user, it won't matter who does the lifting!

_____

- Work intensive

- Can feel 'fake' if not executed convincingly

Example 1
In the 1980's, IBM planned to build a 'revolutionary' speech-to-text machine. To test whether users wanted it, they asked users to talk while a hidden human typed what they said onto a screen. Users were amazed but soon got tired of talking and preferred typing themselves. Using an 'automaton' saved IBM from sinking their entire business into speech-to-text!

Example 2
We worked with a personalised learning organisation in Kenya. Before they invested in developing complex algorithms to personalise content in their app, they offered the app with content to teachers. Teachers worked with children to suggest the next activity to them while they used the app in breaks between lessons. Our partner then spoke to teachers to test whether children stayed engaged with the content, and how it affected their motivation and learning.

2. The Pinocchio

Build a non-functional, 'lifeless' version of the product to see how people respond.

+ Really, really cheap
+ Doesn't need the same technology expertise as building something that really 'works'
+ Gets at user behaviour not just opinion

- Needs some 'suspension of disbelief' from test subjects
- The data gathered can be limited, as users interact with something lifeless

Example 1
In 2018, Onyx Connect put 'fake GPS' units on their pay-as-you-go bicycles in Zambia to test if this deterred bike theft. The GPS units looked convincing but were switched off and not actually collecting location data (which is expensive and requires regulatory approval).

Example 2
We worked with Deaf Reach, a non-governmental organisation (NGO) in Pakistan, providing schools for deaf children. To test whether they might provide laptops with videos to children, they first gave the children laptops and trained the families on charging, maintaining, storing, and basic use. Once they knew the laptops would be looked after, they invested in developing educational videos in sign language.

3. The Minimum Viable Product (MVP)

Create a functional version of the idea but stripped down to its most basic functionality.

_____

+ Allows testing of the most crucial parts of a proposition

+ Much higher fidelity than other testing methods like the Automaton Chess Player or Pinocchio: users will actually use the thing

_____

- You have to be clear about what's crucial to include and what can be included later

Example 1
Worldreader bring digital books to disadvantaged children. In 2010, they launched the MVP of Worldreader mobile to take advantage of increasing dumb-phone ownership in Africa. It had nothing except lists of books and a simple text reader — no book covers, descriptions, ratings, comments, or bookmarks.

Example 2
We worked with an EdTech partner building a telephone helpline connecting out-of-school adolescent girls with teachers. Before investing in a partnership with a telecommunications company, we asked a small group of 30 teachers to ring their existing students for two weeks, testing whether they were able to have conversations about learning.

4. The Provincial

Before launching worldwide, run a test on a very small sample.

—————————————————————————

+ Constrains logistical, cultural, and place-based factors to a manageable scale

+ Cost and resource efficient

+ Helps you experience the entire end-to-end journey of an intervention — maximising chances of discovering things that might surprise you

—————————————————————————

- Can exclude some people from getting the benefits of a new thing

- Saves money but not always time

Example 1
Before rolling out new Internet-of-Things-enabled solar energy systems across health facilities in Zimbabwe, Africa Power Storage (a solar-tech company) tested the end-to-end journey in two locations — a smaller clinic and larger hospital. This let the company learn about shipping, last mile logistics, installation, maintenance, and impact — without making an important financial outlay.

Example 2
To test radio-based models in Uganda during Covid-19-related school closures, we worked in the Lango region of northern Uganda. We were able to work with Mango Tree Literacy Lab, a grassroots organisation who knew the area very well.

5. The Fake Door

Create a fake 'entry' for a product that doesn't yet exist in any form.

_____

+   Really fast, really easy first-stage validation: is there any interest/need?

+   One step up from just asking people (where it's easy to say yes). It involves someone doing something (which is harder to say yes to)

_____

-   Only captures expressed interest, not follow-through

Example 1
In 2007 — and before he'd even built a prototype — Dropbox Founder Drew Houston created a 4-min 'explainer video' to test whether his product solved a problem for people. At the bottom of the video was a form to join the waiting list for the private beta.

Example 2
We are working to identify what EdTech interventions work for the most marginalised learners in Bangladesh. To see if we had a range of (or any) EdTech partners we might work with, we created a GoogleForm. It asked respondents to complete some basic information (a 15-minute task), and to tick a box if they were interested in being involved in our initiative.

6. Time, Money, or Reputation

Don't just ask users or stakeholders if they think it's a good idea — ask them to put their time, money, or reputation on the line.

+ Prevents people from just saying nice things or get you to stop asking more questions

+ Time and reputation are a good way to get from zero to some commitment, which can be built on later

- Money is the only 'real' form of commitment — time and reputation might still give you false positives

Example 1
We were lucky enough to pitch our education helpline idea to a Minister of Education in a country in East Africa. To find out whether she was really interested in supporting our idea in future budgets, we asked her if we could have a dedicated focal point from her team — asking her for their time, and asking her to put her reputation on the line.

7. The Pretend-To-Own

**Before investing in buying whatever you need, rent or borrow it first.**

_____

+ Avoids 'wasted' investment due to pursuing the wrong needs

_____

- Requires rental of equipment/service infrastructure for a sensible amount of time: you still need to give people time to use and respond

- You might need to spend some social or reputation capital (!)

- You may have to change the service if a rented model differs significantly from an owned model

Example 1
Before investing in a GBP 70,000 drone, a humanitarian NGO asked to borrow it from a drone supplier and ran three days of testing in open field in the USA. This validated some of the core features — such as providing a 3G connection and precision dropping — and meant the NGO could ask for modifications from the supplier too.

Example 2
We were working with the government of Zanzibar to develop a nationwide virtual learning environment for secondary school students. Before investing in developing the content repository ourselves, we partnered with an EdTech organisation that already had one, and tested it over a period of one month.
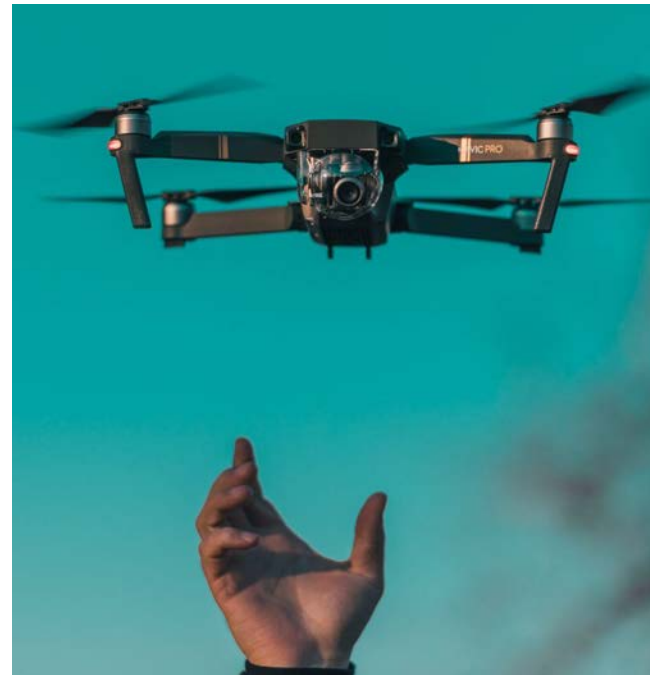
8. The Lean Cost Model

Build a lean cost model to give you an estimate of the cost per year per child. Then, work out how many children you wish to serve, so you know how much money you need to raise or generate.


+   Gives clarity on the commitment needed from users or funders

+   Makes abstract ideas 'real', even before they are implemented

+   Checks an idea passes a basic test (i.e. not way too expensive)

-   Doesn't get users' response to price — only what the price needs to be

-   Is based on hypothetical costs, which could change once you start

Example 1
In the early 2000s, Elon Musk built a financial model to test whether commercial space exploration was viable. He broke down forecast costs and revenue and used basic maths to work out how much a customer would have to pay (or investor would have to fund), and how many customers he needed.

Example 2
We were working with a personalised learning organisation in Malawi. Major donors told us that they would only consider interventions that cost less than USD 10 per year per child. Doing a lean cost model initially indicated that our intervention was almost four times that amount, and showed us where we might reduce the cost.
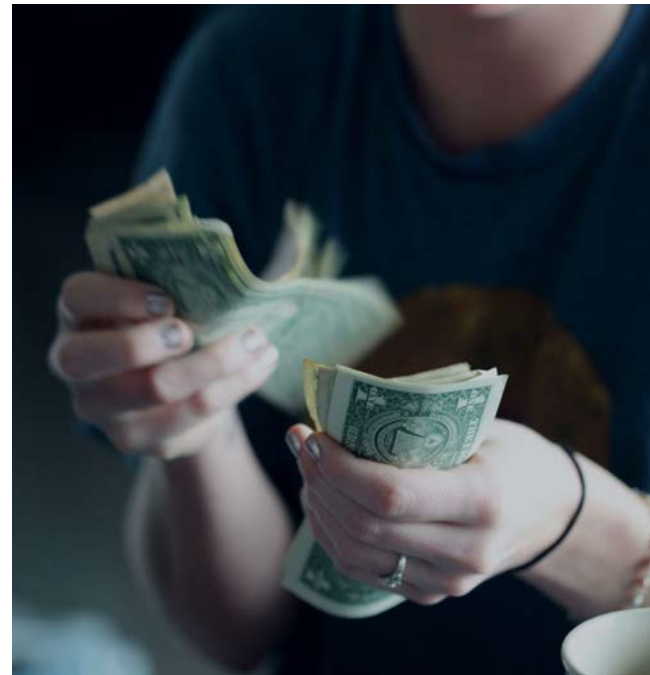
9. The Re-Label

Put a different label on an existing product that looks like the product you want to create.

+ Avoids 'wasted' investment due to something that might not work as you expected

+ Usually gets you 80% of the functionality and value of your product, with 5% of the investment

---

- Means that someone needs to have what you need / do what you want to do

- May distort user expectations of experience or put them off if the 're-label' is not suitable

Example 1
To test whether young people in Kenya would engage with a sexual health and well-being chatbot, the team developing it got some of them to download Roo (a similar chatbot) for a week. Although built for a US (rather than Kenyan) audience, it meant they could get feedback on whether they engaged with the chatbot and found it helpful.

Example 2
When distributing tablets nationwide in Malawi, we were very unsure about whether we could set up a distribution network. To test this, we agreed a Memorandum of Understanding with the government of Malawi to use their existing distribution network for textbooks and other school supplies. Using this network got us to most of the places we needed and allowed us to learn a lot before establishing our own network.

10. The Mini-RCT

Change one thing for half of the users, and track one metric to validate whether that thing has an impact.

+ Let's you systematically isolate and see the impact for one variable (feature)

+ Easier to set up than a multivariate test

- Might seem like one set of users gets unfair benefits that another doesn't

Example 1
The NGO Deaf Reach wanted to see if regular interactions with teachers helped out-of-school children using sign language videos to learn. Having provided 100 children with an offline laptop and pre-loaded content, they provided 50 of that group with a basic smartphone so they could use WhatsApp to call a teacher twice a week.

To assess the impact of the intervention, they tracked how many days per week each group of deaf students engaged with the content. They also administered a basic test to all 100 children at the start and after one month to see if there was any improvement as a result of the additional interaction.

11. The Multivariate Test

Think of a sample of users as a portfolio, measuring how different interventions turn the dial on one or two key metrics.

+ Lets you systematically test a wide range of ideas, making the most of a sample

- Might seem like one set of users gets unfair benefits that another doesn't

Example 1
In Bangladesh, we are working with 20 schools in the Bandarban region. The schools are a mix of government and non-government schools, and in urban and rural areas. We split this sample into 5, with some getting intervention #1 (centred on projectors or TVs), some getting intervention #2 (centred on in-classroom tablets), some getting intervention #3 (centred on out-of-classroom tablets), some getting #1 + #2, and some getting #1 + #3. To keep other salient factors from clouding our results, we made sure that within each group, there were both government / non-government and urban / rural schools. We interviewed teachers and carried out classroom observations across all 20 schools to see which variation was most effective.