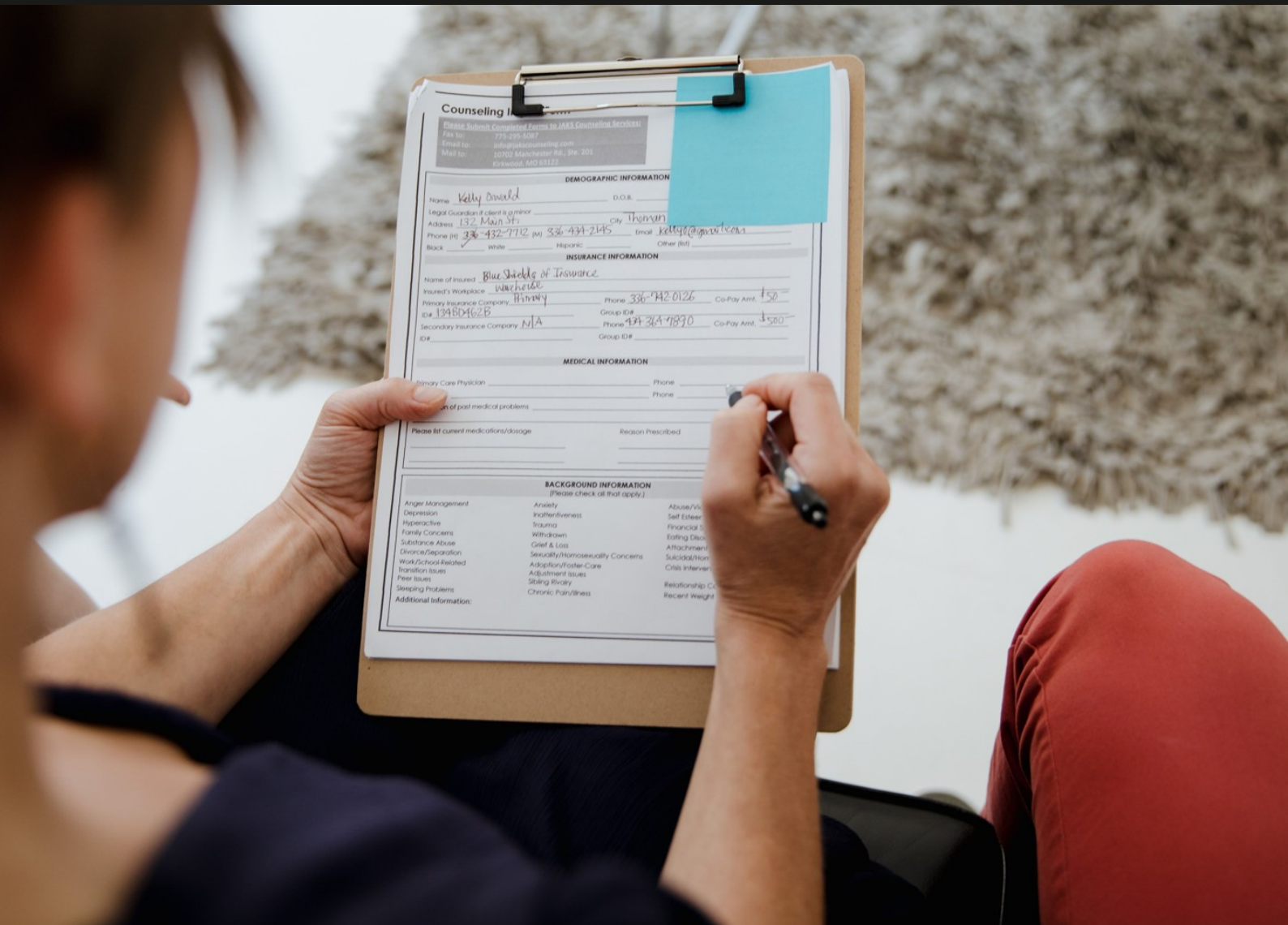




TRANSITIONS LAB • METHODOLOGY

Measuring What Actually Changed

The Lab's method for field research and impact measurement: Scope, Design, Collect, Deliver, read across Reach, Depth, and Experience.

START HERE

How to use this resource

This resource sets out the Lab's method for field research and impact measurement: a four-step process (Scope, Design, Collect, Deliver) and a three-dimensional lens (Reach, Depth, Experience). It explains how the Lab produces evidence of what actually changed, for whom, rather than a count of activities delivered.

What this is

A working reference, not a marketing brochure. Read it in order, or jump to the section you need using the contents opposite.

Who it is for

Funders and programme managers who need to commission or judge impact evidence, organisations preparing to be evaluated, and researchers who want a repeatable, honest measurement discipline.

How to get value from it

- 01** Understand output versus impact
Section 1 draws the distinction the whole method rests on: what was delivered versus what actually changed.
- 02** Follow the four steps
Section 2 walks through Scope, Design, Collect, Deliver, the process every engagement runs.
- 03** Learn the three-dimensional lens
Section 3 explains Reach, Depth, and Experience, and why all three are needed to avoid a misleading picture.
- 04** Use it to commission or judge evidence
Take the framework into your next evaluation as a standard for what good impact evidence looks like.

CONTENTS

What is inside

- 1** The difference between output and impact
Why counting activities is not measuring change.
- 2** The process: Scope, Design, Collect, Deliver
The four steps every engagement runs.
- 3** The measurement lens: Reach, Depth, Experience
Three dimensions, read together, and benchmarked.
- 4** Baseline, and honest measurement
Why a before-picture is non-negotiable.
- 5** The toolkit: choosing the right method
Quantitative, qualitative, and mixed-method instruments.
- 6** The hardest question: causal inference
Separating the change you caused from change anyway.
- 7** Theory of change
The cause-and-effect chain that measurement tests.
- 8** A worked example
The method applied to a market-expansion study.
- 9** Working with the Lab
What a measurement engagement involves.

METHODOLOGY

The difference between output and impact

Most measurement counts what was done. Impact measurement asks what actually changed, for whom, and through what chain of cause and effect. The gap between the two is where most evidence quietly fails.

"Participants reached." "Workshops delivered." "Units sold." These are outputs: easy to count, and they tell you almost nothing about whether a life or an operation actually changed. A programme can report impressive outputs and deliver no impact at all. The reverse is rarer but real: modest activity that changes everything for the people it touches.

This methodology has two halves. The first is a four-step process, Scope, Design, Collect, Deliver, that governs how a study runs. The second is a three-dimensional lens, Reach, Depth, Experience, that governs what a study measures. Together they turn a vague question (did it work?) into a precise, defensible answer.

4

steps from decision to evidence

3

dimensions of real impact

0

impact without a baseline



Measurement begins with people

Every number in a Lab report traces back to a real conversation in a real place. The method exists to make those conversations rigorous, comparable, and honest, so that what a person says actually changed becomes evidence a funder can rely on.

The people who experience an impact know it best. Our instruments are built around what they say matters, not around what is easiest to count.

THE MEASUREMENT METHOD

THE PROCESS

Scope, Design, Collect, Deliver

The process is deliberately simple. The value is in how well each step is done, and in starting from a decision rather than from data. Every engagement runs the same four steps, decision-first, so the study is designed around the question that matters before any instrument is built.

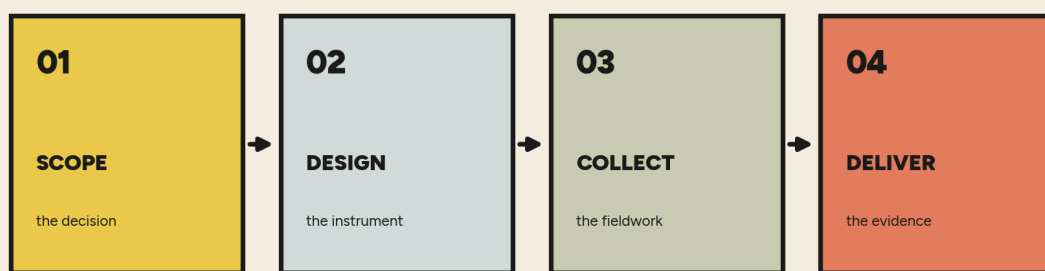


Figure 1. The four-step process. Every engagement runs decision-first: Scope establishes the real question before any instrument is built.

The order is not arbitrary. Each step depends on the one before it: you cannot design an instrument until the decision is clear, cannot collect well until the instrument fits, and cannot deliver anything worth acting on until the collection was done properly. Skip or rush an early step and every later one inherits the flaw. The pages that follow take each step in turn.

THE PROCESS, IN DETAIL

The four steps

01

Scope

Start from the decision, not the data

Every study begins with the decision the client actually faces, not the dataset they imagine they want. We establish the real question, the population that holds the answer, and the evidence that would genuinely change a mind. Most research is wasted here, by gathering data that was never going to settle anything.

02

Design

Build the instrument to fit the question

We design the study to match the question: interviews where depth matters, surveys where comparability matters, baseline-and-endline where change is the claim. This is where technical and social understanding must meet: the instrument has to grasp both the system and the people using it.

03

Collect

Reach the people, properly

Trained researchers who speak the language and know the place reach respondents directly, with informed, recorded consent. In difficult contexts, low-connectivity, multilingual, informal, this step is where most studies quietly fail, and where fieldcraft matters most.

04

Deliver

Evidence you can act on

The client receives clean data, a clear analysis, and a written reading of what it means, benchmarked where possible and honest about limits. Turnaround is measured in weeks. Instruments are built for repetition, so a one-off study becomes a time series.

THE MEASUREMENT LENS

Reach, Depth, Experience

A study measured on one dimension misleads. Count only reach and you learn how many were touched but not whether anything changed. Measure only depth and you miss who was left out. The three together give a complete, honest picture.

Reach.

Who is actually being affected, and, just as important, who is being left out. The easiest dimension to measure and the most often mistaken for the whole story. The critical question is not how many, but who is missing.

Depth.

How materially lives, work, or systems actually change, beyond the headcount. A programme with narrow reach and real depth is often more valuable than one with broad reach and none.

Experience.

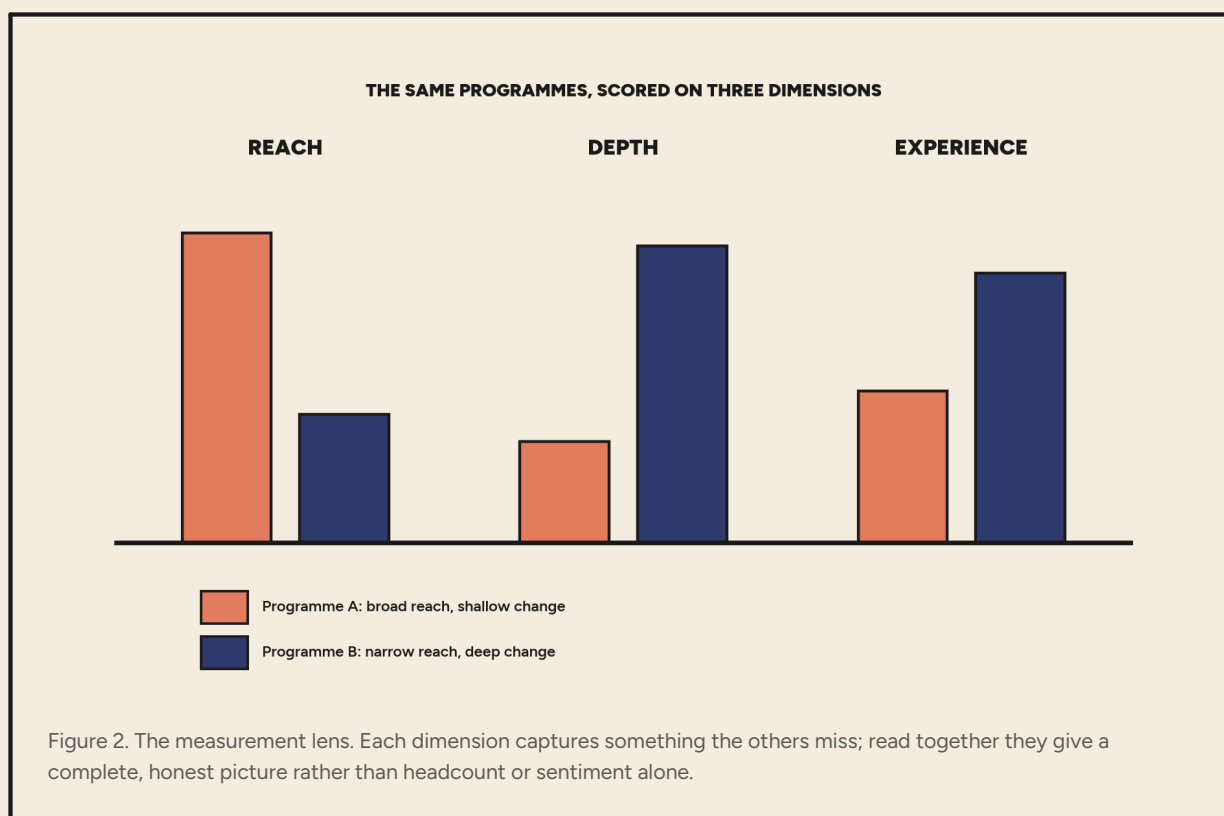
What the change feels like to the people inside it, the friction they meet, and the single thing they would fix first. Where a satisfaction score of seven-out-of-ten is unpacked into the actionable truth the number was hiding.

Dimension	The question it answers	How we measure it
Reach	Who is affected, and who is left out?	Surveys, coverage data, sampling
Depth	How materially did things change?	Baseline-endline, magnitude measures
Experience	What does the change feel like?	In-depth interviews, lived-experience data

THE LENS, ILLUSTRATED

Why one dimension is never enough

The same intervention scored on all three dimensions tells a fuller story than any single one. Consider two programmes: one with broad reach but shallow change, another with narrow reach but deep change. On a headcount alone, the first looks better. Read across all three dimensions, the judgement can reverse.



Programme A (broad reach, shallow change) and Programme B (narrow reach, deep change) are common real patterns. A funder choosing between them needs both dimensions in view, and the experience dimension on top, to know which one actually served its people. Reporting only reach would recommend the wrong one.

READING A RESULT

Performance, benchmarked

A number on its own says little. 82% on access is only meaningful next to what comparable programmes achieve. The Lab reads every dimension against a benchmark, so a client can see not just how a programme performed, but how it performed relative to what was possible.

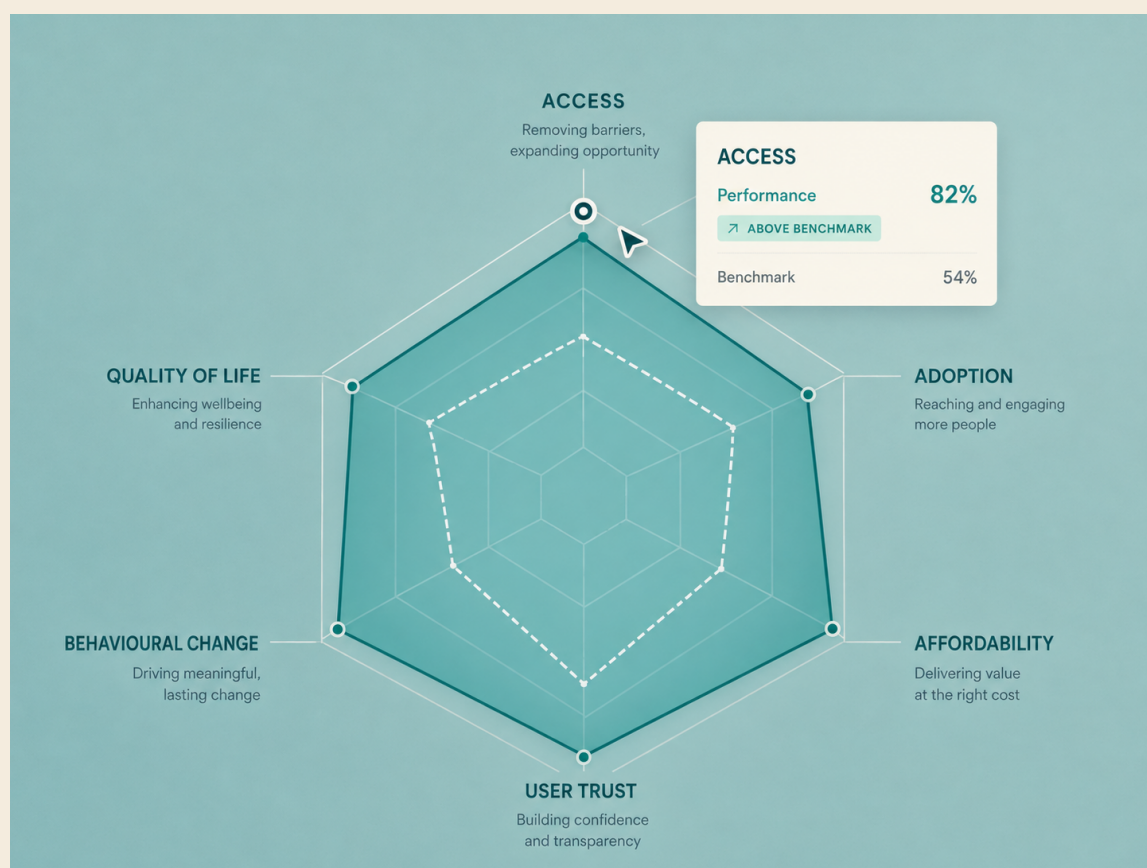


Figure 3. A benchmarked impact profile. Each axis is one dimension of impact, scored and compared against a sector benchmark (the dashed line). The shape shows at a glance where a programme is strong, where it lags, and where it is genuinely above the field.

The value of a profile like this is that it resists the single headline number. A programme can be far above benchmark on access and below it on trust, and both facts matter. Reading the whole shape, rather than one axis, is what turns a score into a decision.

THE DISCIPLINE THAT DECIDES EVERYTHING

No baseline, no impact

Impact is a change, and a change needs a before. This single discipline separates real impact measurement from after-the-fact assertion, and it is the one most often skipped.

Measurement is too often designed near the end of a project, once there is a report to write. By then the before can only be reconstructed from memory and rounded numbers, which is to say it cannot be reconstructed at all. Without a baseline captured at the start, using the same instrument that will be used at the end, every impact claim rests on assertion.

Capture the baseline first

The before must exist because it was recorded, not because someone remembers it. Same instrument at endline. Data stored and version-controlled. Non-negotiable.

Measure while it is live

Midline readings turn measurement from a post-mortem into a steering instrument, catching what works while there is still time to act. Evidence that arrives after every decision has no value beyond the archive.

KEY POINT

The honesty check

Every Lab study ends with questions most reports skip: did the change happen because of the intervention, or alongside it? What else could explain the result? Who was left out, and does that change the headline? Writing down these answers is the difference between an evaluation and a press release.

THE TOOLKIT

Choosing the right method

There is no single best method, only the method that fits the question. A good study often combines several. This section introduces the main methods the Lab uses, and when each earns its place.

The first choice is between quantitative methods, which measure how much and how many across a large group, and qualitative methods, which explain how and why in depth. Most of the Lab's work is deliberately mixed-method: numbers establish the pattern, and depth explains it.

Family	Answers	Trades off
Quantitative	How many, how much, how often. Comparable across people and time.	Depth. A number rarely explains itself.
Qualitative	How, why, and what it means to the people involved.	Scale and comparability. Harder to generalise.
Mixed-method	Both, with each covering the other's blind spot.	Cost and design effort. Worth it when the stakes are real.

The pages that follow describe the specific instruments within each family, and the design choices, sampling, comparison, causal inference, that decide whether the evidence holds up.

QUANTITATIVE INSTRUMENTS

Measuring at scale

Quantitative methods establish the pattern: how widespread an effect is, how large, and for whom. Their strength is comparability; their discipline is sampling.

Instrument	What it is, and when to use it
Structured survey	A fixed set of questions asked identically of many people. The workhorse of reach and magnitude measurement. Only as good as its sample and its questions.
Baseline-endline	The same survey run before and after, so change can be measured rather than asserted. The backbone of credible impact claims.
Administrative data	Records a programme already generates (sales, usage, attendance). Cheap and continuous, but rarely designed to answer the impact question directly.
Sensor and IoT data	Direct measurement from devices, real-time usage, movement, energy. Powerful where behaviour leaves a digital trace, as in the Lab's mobility work.

KEY POINT

The sampling discipline

A survey's credibility rests entirely on who was asked. A biased sample produces confident, precise, wrong answers. The Lab designs samples to represent the population that holds the answer, including the hard-to-reach who are most often left out, and states the sample's limits plainly.

QUALITATIVE INSTRUMENTS

Understanding in depth

Qualitative methods explain what the numbers cannot: the reasoning, the friction, the meaning. They are how the Lab reaches the depth and experience dimensions.

Instrument	What it is, and when to use it
In-depth interview	A guided one-to-one conversation that follows the participant's reasoning. The richest source of why, and of the detail a survey cannot anticipate.
Focus group	A facilitated group discussion that surfaces norms, disagreements, and shared language. Strong for social dynamics; weak where privacy or power make people guarded.
Field observation	Watching what people actually do, rather than what they report. Catches the gap between stated and real behaviour that self-report misses.
Ethnographic immersion	Extended presence in a setting, building the contextual understanding that short visits cannot. Reserved for questions where context is everything.

The Lab's interview design follows a deliberate depth ladder: from fact (what happened), to behaviour (what you did), to reasoning (why), to meaning (what it changed for you). A survey typically reaches only the first rung; the actionable insight usually lives on the third and fourth.

THE HARDEST QUESTION

Did the intervention cause the change?

The central difficulty in impact measurement is attribution: separating the change an intervention caused from change that would have happened anyway. Getting this wrong is how programmes claim credit for trends they did not create.

The gold standard is the counterfactual: what would have happened without the intervention. Because that world cannot be observed directly, it must be approximated. The main approaches, in descending order of rigour and cost:

Approach	How it approximates the counterfactual
Randomised control trial	Randomly assign who receives the intervention; the untreated group shows what would have happened. Rigorous, but often costly or ethically impossible.
Quasi-experimental	Construct a comparison group that resembles the treated one (matching, difference-in-differences). Strong where randomisation is not possible.
Contribution analysis	Trace the causal chain and test it against evidence, asking whether the story holds and what else could explain the result. Suited to complex, real-world settings.
Pre-post only	Compare before and after with no comparison group. The weakest basis for attribution, and honest reporting says so.

The Lab matches the approach to the stakes and the setting, and is explicit about the strength of the claim the method can support. An honest ‘this is associated with’ is worth more than a false ‘this caused’.

PUTTING IT TOGETHER

Designing the mix

A real study is a designed combination, not a single instrument. The art is sequencing the methods so each answers what it is best at and covers another's blind spot.

Quant then qual

A survey finds that adoption dropped in one region; interviews then explain why. The numbers locate the question; the depth answers it. The Lab's most common design.

Qual then quant

Interviews surface an unexpected factor; a survey then measures how widespread it is. Used when the important variables are not known in advance.

KEY POINT

Triangulation

Confidence comes from agreement across methods. When a survey, interviews, and observation all point the same way, a finding is robust. When they disagree, that disagreement is itself a finding, and usually the most important one. The Lab treats convergence, not any single number, as the test of a result.

Whichever methods are chosen, they serve the same end: evidence that is comparable where it needs to be, deep where it needs to be, and honest about what it can and cannot conclude.

BEFORE MEASUREMENT

What a theory of change is

Before anything can be measured, there has to be a claim about how change is supposed to happen. That claim is a theory of change: the explicit chain of cause and effect from what a programme does to the difference it intends to make.

A theory of change makes a programme's assumptions visible and testable. Without one, measurement has nothing to test: it collects numbers with no way of knowing which ones would confirm or refute that the programme is working. With one, every link in the chain becomes a hypothesis that evidence can check.

The standard chain runs through five links, each a step further from the programme's control and closer to the change that matters:

Link	What it is	Example
Inputs	The resources put in.	Funding, staff, vehicles.
Activities	What the programme does.	Deploying electric bikes, training riders.
Outputs	The direct, countable products.	Bikes on the road; riders trained.
Outcomes	The changes in behaviour or condition.	Riders switch; incomes rise.
Impact	The lasting difference, net of what would have happened anyway.	Durable shift in livelihoods and emissions.

Most measurement stops at outputs, because outputs are easy to count and fully within the programme's control. But outputs are not impact. The theory of change is what forces the harder question of whether the outcomes and impact actually followed.

THEORY OF CHANGE, CONTINUED

Making it testable

A theory of change is only useful if it can be wrong. The Lab builds one at the start of an engagement, and treats each link as a hypothesis with its own evidence, its own risks, and its own measurement.

Two things turn a diagram into a working instrument:

- › Assumptions made explicit. Every arrow between links hides an assumption, that trained riders will actually switch, that switching will actually raise income. Naming these assumptions shows exactly where a programme's logic is most likely to break, and therefore where to look hardest.
- › Indicators attached to each link. Each link gets a measurable indicator and a baseline, so the chain can be tested step by step. When impact does not appear, the indicators show which link broke, whether the activity did not happen, the outcome did not follow, or the impact was captured elsewhere.

KEY POINT**Why the Lab starts here**

A theory of change built at the proposal stage does double duty: it strengthens the design by exposing weak assumptions before money is spent, and it defines exactly what to measure so the evaluation is not improvised at the end. It is also what EU funders increasingly expect, and what the Lab's European Impact Tracking work is built around.

A programme with a clear, tested theory of change can say not just that it worked, but why, and can be improved link by link. That is the difference between an evaluation that reassures and one that teaches.

A WORKED EXAMPLE

The method applied: a market-expansion study

To show the method in use, consider a manufacturer weighing whether to expand electric two-wheelers into a new region. The four steps and three dimensions turn a broad question into evidence a board can act on.

**WORKED
EXAMPLE**

Should we scale into this market?

Scope. The real decision was not "are riders satisfied" but "does the daily economics work well enough for adoption to hold once the subsidy ends?" We identified the rider segments that would answer it.

Design. A short, repeatable survey on running costs and switching intent (for reach and depth), paired with a smaller set of in-depth rider interviews (for experience). A baseline was captured before any expansion.

Collect. Local researchers who ride the same routes reached riders directly, in their own language, with recorded consent, catching the segments a call-centre survey would have missed.

Deliver. Within weeks: a segmented read on where adoption was real, where it was fragile, and what would have to change to hold it, plus the instrument to run the same measurement again next quarter.

Read across the three dimensions. Reach showed who was adopting and who was priced out; depth showed how much running-cost savings actually moved behaviour; experience surfaced the per-transaction fee riders would fix first, the finding a satisfaction score had hidden.

One question, answered in a way a decision could rest on, and repeatable, so the one-off study became the first point in a trend.

WHY THIS METHOD, AND WHY THE LAB

Rigour, independence, usefulness

This method is built to avoid the two ways impact measurement usually fails: too elaborate to repeat affordably, or too thin to tell you anything you did not already assume. The Lab's standard is the middle ground, structured enough to compare, rich enough to matter.

Independence is the product.

A programme measuring its own impact has every incentive to find success. That is not dishonesty; it is gravity. Independent measurement exists to counter it.

Built for repetition.

Because our instruments are standardised, a single study becomes a trajectory that can be managed, compared, and defended over time.

Written to be used.

Findings are returned in weeks, in language a decision-maker can act on without a translator, tied to the choice in front of them.

Rigour that no one can use has failed at the last step. We write evidence to be acted on.

THE MEASUREMENT METHOD

WORKING WITH TRANSITIONS LAB

From method to measurement

This resource sets out how the Lab measures impact. A live study draws on further method, matched to your decision and the context it sits in.

IN THIS FREE RESOURCE	IN AN ENGAGEMENT, WE GO FURTHER
The four-step process and the reach, depth, experience lens.	› Instruments designed to your decision, with sampling and question design built for your population.
The baseline discipline, explained.	› Baseline, midline, and endline fielded and version-controlled, so change is evidenced, not asserted.
The honesty check, as a principle.	› Formal contribution analysis: what else could explain the result, and who was left out, tested rather than noted.
A worked market-expansion example.	› Benchmarking against comparable studies, so your numbers have a reference point beyond themselves.

The method here is the shape of the work; an engagement is where it is fitted to your question and run in the field.

If a decision in your work turns on evidence of what a technology or programme actually does to real people, we are glad to talk it through, with no obligation. transitionslab.org

Transitions Lab is an independent research team that studies how technologies and programmes meet real people, social science first, technical literacy alongside. This methodology may be used and cited with attribution. transitionslab.org