SIDESTREAM · WHITEPAPER

The 4-Level Measurement Framework

How to prove your behaviour change programme actually worked.

BEN LAUMANN · CO-FOUNDER, SIDESTREAM

Behaviour Change Consultancy · London & Germany

sidestream-consulting.co.uk

THE PROBLEM IN ONE SENTENCE

Most L&D teams cannot prove their behaviour change training works.

Ask most L&D leaders how they measure the impact of their behaviour change programmes and the answer comes back in two parts: a satisfaction score, and a story. The first is data nobody trusts. The second is data nobody can scale. Meanwhile the CFO is waiting for an answer to a different question: **did the behaviour actually change, and what did that change earn us?**

This whitepaper walks through the four-level framework Sidestream uses to answer that question, and the three numbers your finance team will actually accept as ROI evidence.

22%	70%	£1,068
Of orgs measure beyond Level 2	Of change programmes fail	Avg UK L&D spend / employee

Sources: Kirkpatrick Foundation; McKinsey Transformation Research; CIPD Learning at Work, 2024.

THE FRAMEWORK

Kirkpatrick's Four Levels

Donald Kirkpatrick's 1959 model is still the gold standard for training evaluation. It defines four progressive levels of impact. Most L&D reporting stops at Levels 1 and 2, the levels that are easiest to measure but tell you nothing about ROI. Levels 3 and 4 are where behaviour and business outcomes live.

LEVEL 01

REACTION

Did participants enjoy it?

Satisfaction scores at the end of the programme. Easy to measure, almost impossible to act on. Tells you whether the room felt good, not whether anything changed.

LEVEL 02

LEARNING

What did they learn?

Knowledge checks, post-course quizzes. Measures retention immediately, rarely 30 days later. The classic "passes the module, fails on the job" gap.

LEVEL 03

BEHAVIOUR

Are they doing it differently?

On-the-job behaviour 60–90 days after the programme. Requires baseline + post-measurement with the same instrument. This is where ROI begins.

LEVEL 04

RESULTS

Did the business outcome shift?

The number that mattered before the programme (turnover, engagement, profit, time-to-promotion) actually moved against baseline. This is what finance signs off on.

THE PROBLEM

Why Most L&D Stops at Level 2

Research from the Kirkpatrick foundation shows that **only 22% of organisations measure beyond Level 2**. That is a problem because **only Levels 3 and 4 predict ROI**. Levels 1 and 2 measure how the room felt and what people remembered the next morning; neither correlates with on-the-job change.

> The blunt truth: if your training reporting tops out at "92% satisfaction" and "post-course quiz score 84%", you have no evidence anyone is doing anything differently. You have evidence of an enjoyable few hours.

Three reasons it happens

1. It is what the supplier is allowed to ask. Suppliers measure what looks good in their own report. Reaction is the metric most likely to flatter them.

2. Levels 3 and 4 require time and trust. A 90-day re-measure means the supplier is still around 90 days later; most are not.

3. Nobody agreed the metric in advance. Without a baseline number set in week 0, "ROI" is back-fitted from whatever data happens to be available.

THE EVIDENCE

The Three Numbers Your CFO Actually Wants

For finance to take an L&D programme seriously as an investment rather than a cost, it needs three pieces of evidence. Get these right and the conversation changes.

1. The Behavioural Delta

The difference between a measured baseline behaviour **before** the programme and the same measure 90 days **after**. The key word is *same*: the same psychometric tool, the same observation framework, the same population.

Tools we use: Team Diagnostic Survey · Edmondson Psychological Safety Index · Leadership Practices Inventory (LPI) · behavioural-coding of recorded interactions.

2. The Cost of the Old Behaviour

Behaviour change becomes ROI when you can attach a cost to the behaviour you replaced. Examples:

· **Disengagement** → **Turnover cost.** Replacing a mid-level employee costs a significant share of annual salary. Use your own HR figure rather than a rule of thumb.

· **Manager defaults** → **Team performance.** Gallup attributes 70% of variance in team engagement to the manager. Engagement uplift of one quartile correlates with **21% higher profitability**.

· **Failed transformations** → **Re-do cost.** Catching resistance in a simulation before launch protects milestone dates. Use your own cost per month of delay.

3. The Programme Cost (All In)

Total Cost of Ownership matters more than headline price. Add: design, delivery, participant time-out, opportunity cost of senior sponsorship, follow-up coaching, and measurement. The true figure is materially higher than the supplier invoice, so add these lines explicitly rather than applying a multiplier.

> ROI = (Cost of Old Behaviour – Cost of Programme) ÷ Cost of Programme × 100.
> We do not publish a typical return figure, because it depends entirely on the cost of the behaviour you are replacing. Work it out with your own numbers.

THE METHOD

How Measurement Gets Built In

This is where most consultancies lose credibility: they offer to "measure later" or rely on the satisfaction survey at the end. Real measurement gets baked into the engagement structure from week 1.

Week 0: Baseline. Agree the behavioural metric and run the validated diagnostic. This is the number you compare against later.

Week 1–3: Audit. Interviews and focus groups give qualitative context. Without this, the numbers are uninterpretable.

Week 9–12: Delivery. The intervention itself, but with deliberate behavioural mechanisms designed to move the metric, not just the mood.

Week 13–24: Re-measurement. 90 days post-intervention, the same diagnostic is rerun. The delta is your evidence.

What measurement is not

· Self-reported behaviour change is not behavioural evidence. People are bad at rating their own behaviour, especially right after a positive experience.

· "% would recommend" is a Net Promoter Score for the supplier, not an outcome metric for your organisation.

· Single-source data (only HR, only managers, only the cohort) is biased. Triangulate.

· Measurement after a single workshop is rarely meaningful. Behavioural change takes 60–90 days to consolidate.

THE WORKSHEET

A Measurement Plan You Can Actually Run

Print this page. Fill it in for one programme, current or planned. If you cannot fill in even half of it, the programme has a measurement problem, not an execution problem.

STEP 1 · DEFINE THE BEHAVIOURAL METRIC (Level 3)

- The specific behaviour we want to change is: ____________________________
- The instrument we will use to measure it: ___________________________ (e.g. Edmondson Index, LPI, custom 360)
- The cohort we will measure: ____________ people in ____________ team(s)
- Baseline date: ____________ · Re-measure date: ____________ ($\Delta \geq 60$ days)

STEP 2 · DEFINE THE BUSINESS METRIC (Level 4)

- The business outcome we expect to move: _____________________________
- Current 12-month baseline of that metric: ___________________________
- Target 12-month change attributable to programme: ___________________
- Owner of the metric (signs off the report): _________________________

STEP 3 · COST OF OLD BEHAVIOUR

- Annual cost we estimate the current behaviour is producing: £________
- Source / calculation method: ____
- Validated with finance? ■ yes ■ not yet

STEP 4 · ALL-IN PROGRAMME COST

- Supplier invoice: £________ · Internal time-out: £________
- Sponsorship & travel: £________ · Measurement: £________
- Total all-in: £________ → Break-even ratio = Total ÷ Cost of Old Behaviour (must be below 1)

SIDESTREAM

Behaviour change you can actually measure.

Sidestream is a UK-based behaviour change consultancy combining organisational psychology, systemic coaching and immersive theatre to design programmes that move Levels 3 and 4, not just Levels 1 and 2. Award-winning work for the Metropolitan Police, the Innocence Project, and HM Courts & Tribunals; founders studied at Cambridge, UCL, Goldsmiths and Bocconi. Our own controlled study was run at UCL in 2020 (n = 57): sidestream-consulting.co.uk/research

Where to next

1. **Run this worksheet** for one programme you currently sponsor. The gaps will tell you more than any audit could.

2. **Book a free 30-minute diagnostic call** at sidestream-consulting.co.uk/contact, and we will help you scope the metric, the baseline and the proof.

3. **Read further:** our five research frameworks (Kirkpatrick, COM-B, Kolb, Ericsson, Roediger & Karpicke) at sidestream-consulting.co.uk/approach.

Ben Laumann · Co-Founder & Managing Director UK
Organisational psychologist · theatre director · studied at UCL, Cambridge and Bocconi
info.sidestream@gmail.com · +44 7542 401088 · sidestream-consulting.co.uk