sopact

SOPACT INTELLIGENCE LIBRARY
BOOK 06 · USE-CASE GUIDE

PRE → MID → POST

sopact

4 CHAPTERS · 21 SECTIONS

FROM LEARNER TO LEARNING INTELLIGENCE.

Unmesh Sheth · Founder & CEO, Sopact · 2026

sopact.com

BOOK 06

CONTENTS

What’s inside.

Four chapters, one persistent learner record. The arc follows a single learner across the only timeline that proves a program worked — **pre**, **mid**, and **post** — and shows why training activity is giving way to learning **intelligence**.

PRE → MID → POST → PROOF

PURPOSE · WHO THIS IS FOR

For the teams who have to prove the training worked.

If you run a workforce program, a corporate L&D function, a fellowship, or an accelerator — and a funder or a leadership team asks not whether learners liked the program but whether they **changed** — this guide is for you. U.S. organizations spent over $100B on training in a recent year, and research found only 12% of employees apply new skills on the job. That 88% gap is not a curriculum problem. It is a measurement-architecture problem.

The argument runs through every chapter: effectiveness shows up 90 days later, in whether a learner applies what they learned. The data exists — in an LMS, a survey tool, an email thread, an HRIS — but no shared learner identity connects it. Learning intelligence fixes that with one persistent record per learner, so the only timeline that proves a program worked — pre, mid, post — stays intact.

1

The shift

From training activity to learning intelligence — and the Learner Identity Break that keeps most programs at Level 2.

2

Pre

Design the instruments before the content — a baseline written in observable, behavioral terms.

3

Mid

Formative checks, real pre/post deltas, and risk flags that surface a struggling learner while the cohort is still running.

4

Post

Transfer at 30–90 days, business results, and a funder report generated in minutes, not weeks.

i

One learner ID runs the whole arc

A persistent ID issued at enrollment carries through every instrument — intake, pre-test, weekly pulse, post-test, mentor observation, and the 90-day follow-up — so the cascade holds instead of fragmenting into four disconnected events.

CHAPTER ONE · THE SHIFT

Effective training is a claim. Measured training is evidence.

Completion rates and satisfaction scores describe activity. Whether a learner applied the skill ninety days later is the evidence — and it only exists if one record connects the learner across time.

THE 88% GAP

Most training produces activity, not outcomes.

Training effectiveness does not show up in completion rates or satisfaction surveys. It shows up 90 days later, when a participant either applies what they learned or does not. That data exists in most organizations — scattered across forms, an LMS, email, and an HRIS, with no thread connecting the training event to the outcome it was supposed to produce.

$100B+
spent on U.S. workplace training in a single year

12%
of employees apply the new skills on the job

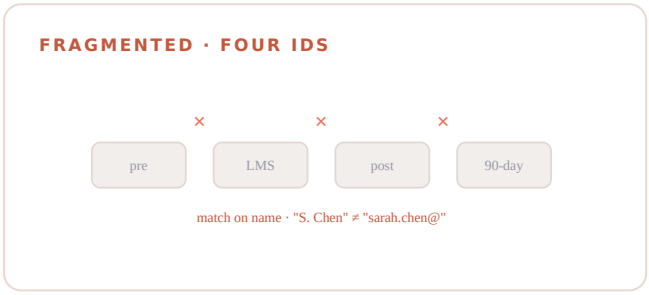
80%
of analyst time spent on cleanup, not analysis

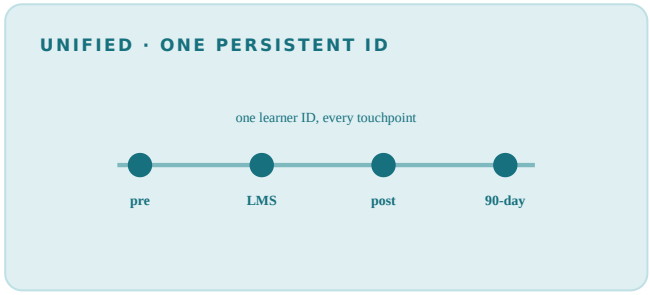
Three things close the gap — none of them a new framework
A pre-training baseline; longitudinal tracking of the same individuals across 30–90 days; and a persistent participant record that survives long enough to correlate learning with performance. Not a better model. Infrastructure.

THE LEARNER IDENTITY BREAK

Four tools, four IDs, no shared learner.

Every tool in the standard stack — LMS, survey platform, follow-up emailer, HRIS — issues its own participant ID. None are shared. When a learner moves between tools, the record terminates and a new one begins. That is the break that makes Level 3 and 4 measurement a manual reconciliation project.





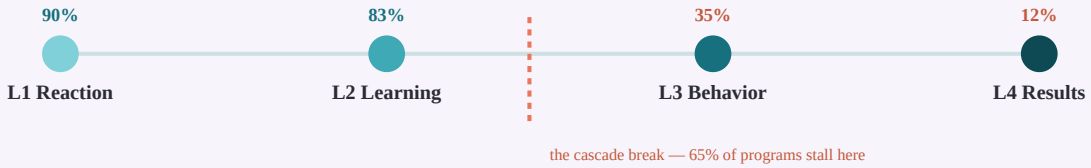
It is an architecture gap, not a feature gap

Level 1–2 (satisfaction, quiz scores) don’t need longitudinal tracking, so traditional tools handle them fine. They break at Level 3+ because measuring behavior change requires connecting the same person’s data across time. A persistent learner ID, assigned at intake, closes the break.

REACTION → LEARNING → BEHAVIOR → RESULTS

The Kirkpatrick cascade — and where it snaps.

The four levels are a cascade: each level's evidence depends on the one before. The model is not broken; the data architecture beneath it is. Break the chain between Level 2 and Level 3, and behavior and results become narrative claims rather than measurements.



L1 REACTION · L2 LEARNING

Did they value it; did knowledge increase. Easy to collect, weakly predictive — and where most programs stop.

L3 BEHAVIOR · L4 RESULTS

Did the skill reach the job; did outcomes improve. The levels funders renew on — reachable only when the ID chain holds.

PRE · BEFORE THE FIRST SESSION

The instruments come before the content.

Design the assessment after the curriculum and it measures what was delivered. Design it before, and it measures whether the learner can now do something they could not do before. That sequence is the whole game.

CONFIRMATION, OR EVIDENCE

When assessment is built after content, it measures delivery.

The assessment question usually arrives last — three weeks before the cohort, after the content is locked. That sequence produces instruments calibrated to what was taught, not to whether learners met the objective. Fewer than 30% of organizations define learning objectives before designing assessment.

CONTENT FIRST **Measures delivery**	INSTRUMENTS FIRST **Measures learning**
Pre-test and post-test written by different people, different formats — gain score is invalid.	**Identical items and anchors** at pre and post — the delta is real.
Rubric drafted after delivery, reflecting observed struggles.	**Rubric anchored to the objective,** written before the first slide.
Follow-up added six weeks later, unlinked — response rate under 20%.	**Follow-up designed up front,** tied to the record — response rate above 60%.

→ **Reverse the sequence**

Define what “learned” looks like in observable, behavioral terms. Then design the baseline, the summative, and the 30/60/90-day follow-up together, with matched anchors and one shared ID — before the curriculum is written. It needs no new technology, only sequencing.

THE STARTING REFERENCE

Without a baseline, a post-score is uninterpretable.

Pre-training assessment establishes each learner's starting point on the exact competencies the program develops — the structural prerequisite for every gain score and effectiveness claim that follows. It has three parts.

1

Knowledge test

The core concepts the program addresses — the same items, scored the same way, at post.

2

Confidence scale

Self-rated ability on each target competency — self-efficacy predicts transfer better than satisfaction.

3

Behavioral baseline

"Describe how you currently handle [scenario]" — the richest comparison point at follow-up.

i

Objectives in behavioral terms

"Employees will be better at customer service" is not measurable. "Customer satisfaction for trained employees will rise 10% within 90 days" is. Vague objectives produce unmeasurable instruments; observable behaviors anchor every instrument downstream.

DESIGNED TOGETHER, NOT IN SEQUENCE

One instrument set, built before the cohort.

Parallel instrument design is the solution to the Backwards Design Gap — but it only works if the infrastructure connects every instrument to the same person across time. Build the whole set at once, with matched competency anchors.

NEEDS ASSESSMENT

Organizational, task, and individual levels — output is a measurable learning objective per competency gap, not a topic list.

BASELINE · FORMATIVE · SUMMATIVE

Pre-training baseline, session-period comprehension checks, and the post-training summative — same anchors, one record.

FOLLOW-UP TRIGGERS

30-day behavioral self-report, 60-day manager observation, 90-day retention test — configured before the cohort completes, not after learners disengage.

✓ **Observer roles, mapped first**

Who assesses whom — facilitator, manager, mentor, peer — each gets a different instrument routed through the same learner record, so triangulated evidence assembles itself.

MID · WHILE THE COHORT IS LIVE

The window to fix it is still open.

Formative data exists to course-correct, not to grade. A struggling learner, a confusing module, a confidence dip — caught in week three, while there is still time to act, not in a report after graduation.

COURSE-CORRECT MID-FLIGHT

Catch the gap in week three, not in the final report.

Formative assessment happens during delivery. Its purpose is correction: knowledge checks, scenario questions, and confidence pulses tell instructors which concepts are not landing in real time — so the next session changes before the cohort graduates with persistent gaps.

SCENARIO CHECKS

3–5 applied questions per content block — testing understanding, not memorization.

CONFIDENCE PULSE

Self-rating before and after each block — the delta predicts which areas need reinforcement at follow-up.

PAIRED-PRACTICE NOTES

Facilitator observation against a shared rubric — structured, not free-text, so it aggregates.

→ **A leading indicator, not a lagging one**

When weekly checks reveal 70% struggled with Module 3, the program lead revises it before the next session — not after the cohort is gone. Formative data is the only kind collected early enough to change the outcome it measures.

LEVEL 2, DONE RIGHT

Knowledge and confidence are not the same signal.

Level 2 requires a paired pre-and-post assessment using identical items, with the delta calculated per individual — not an average across two separate groups. And it pays to score knowledge and confidence separately: they diverge, and the divergence tells you what to do next.

HIGH KNOWLEDGE, LOW CONFIDENCE

The learner knows it but won't apply it yet. The intervention is coaching, not re-training — a different cost and a different fix.

HIGH CONFIDENCE, LOW KNOWLEDGE

The risk case: a learner who will act without the competency. Surface it before they reach the job.

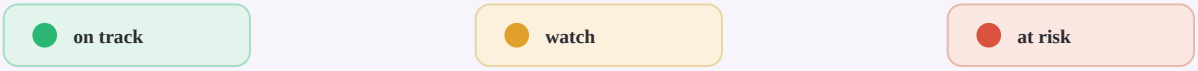
Bound at collection, not after

Quantitative test score and AI-extracted confidence from open reflections live in the same record — so "do high scores predict high confidence?" is visible the moment the last response lands, reproducibly, not a month of analyst work later.

GREEN / YELLOW / RED, EVERY WEEK

See the learner at risk before they drop out.

A connected record produces weekly engagement signals per participant — visible to coordinators during the cohort, not six weeks after it ends. Completion is not engagement; a module clicked through is not a module learned.



Two non-binary indicators beat one completion checkbox: time-on-task, reflection quality, or weekly self-reported barriers.

✓ **Personalized follow-up, 3× the response**

Follow-up surveys delivered as personalized links tied to the original record — not bulk email blasts — produce roughly three times the response rate, which is what keeps the mid-to-post chain intact.

POST · 30, 60, 90 DAYS OUT

Did the skill reach the job?

This is where training effectiveness separates from training evaluation — and where the funder renews or walks. The 30-day mark is critical: skills not applied within 30 days are unlikely to be applied at all.

BEHAVIOR CHANGE ON THE JOB

Ask for the example, not the intent.

The single most important design choice in transfer assessment is behavioral specificity. “Have you used what you learned?” invites a vague yes and overclaims application by two-to-four times. “Describe a specific situation in the past four weeks where you applied [skill]” filters out intent and surfaces evidence.

APPLICATION INTENT **Overclaims 2–4×**	BEHAVIORAL SPECIFICITY **Surfaces evidence**
“Have you used the negotiation framework?” → yes (social-desirability bias).	**“Describe a negotiation in the past month where you used it.”**
Self-report only — no corroboration.	**Learner self-report + manager observation** against the same named behaviors.
Free-text manager emails that can’t be aggregated.	**Structured rubric, AI-scored** — themed across the whole cohort.

Name two to four observable behaviors first

Written as actions (“conducts structured weekly 1:1s”), not qualities (“is a better leader”). Baseline them at enrollment, re-measure at 30/60/90 days from learner and manager, and disaggregate by cohort and role — all on the same ID.

RESULTS, ROI, AND THE FRAMEWORK FIT

Directional evidence beats no evidence.

Level 4 links training to organizational outcomes — productivity, retention, error rates, revenue. You don’t need experimental rigor; you need credible, directional evidence. “Trained teams outperformed by 15%” moves a renewal decision in a way a 4.6/5 satisfaction score never will.

FRAMEWORK	ADDS	USE WHEN
Kirkpatrick (4 levels)	The spine: Reaction→Learning→Behavior→Results	The default for workforce & L&D
Phillips ROI (L5)	Monetizes Level 4 — (Benefits − Costs) ÷ Costs	A CFO is in the funding conversation
Brinkerhoff SCM	Top/bottom performer interviews — narrative depth	The funder wants the story behind the numbers
CIRO / CIPP	Front-loads design quality before outcomes	Multi-phase public-sector programs

→ **Reverse-design from Level 4**

Name the organizational result first, then the behaviors that produce it, then the knowledge those behaviors need, then the experience that delivers it. Measure forward L1→L4; design backward — so Level 3 and 4 evidence is built in, not retrofitted. All four frameworks need the same persistent-ID architecture underneath.

GENERATED, NOT ASSEMBLED

A funder answer in four minutes, not six weeks.

When pre, mid, and post share one learner ID, the evaluation report is a query against the accumulated record — metrics and narrative together, as a live link that updates as data arrives. The reconciliation that used to consume the cycle never has to happen.

PRE/POST DELTAS

Cohort and per-learner gain, disaggregated by segment defined at intake.

BEHAVIOR-CHANGE EVIDENCE

AI-extracted themes from mentor and manager notes, linked to each learner.

ENGAGEMENT DASHBOARD

Weekly risk flags, visible during the cohort.

FOLLOW-UP COMPLETION

30/60/90-day response, 3× the unlinked rate.

FUNDER NARRATIVE

Metrics + stories in one shareable live link.

MULTI-COHORT ARCHIVE

Reused instruments — Year 2 compares to Year 1 automatically.

✓ **Worked example · a mentorship cohort**

A real virtual mentorship program — 60 participants, 6 mastery skills. Intake, weekly check-ins, pre/post, and mentor observation all run on one persistent learner ID, with Level 3 evidence from mentor rubrics. A question that took three days now answers in four minutes — and analysis time per cohort drops from ~200 hours to ~20.

THE THROUGH-LINE

Pre, mid, post — one learner.

The arc of this book is the arc of a single learner ID. Baseline before the first session, formative correction while the cohort runs, and transfer ninety days after — all on one record. Nothing resets at the boundaries, which is the only reason Level 3 and 4 ever hold.

MEASUREMENT

Identical instruments at pre and post, behavioral specificity at follow-up, knowledge and confidence scored apart.

MANAGEMENT

Mid-flight correction, risk flags during the cohort, and a funder report that holds up to Level 3–4 scrutiny.

SEE IT ON YOUR LAST COHORT

Bring an intake form. Map the cascade live.

Bring your current intake form and your last cohort’s data — even if it’s scattered across Google Forms, an LMS, and a spreadsheet. We map your Kirkpatrick architecture on your real program: persistent IDs, named behaviors, mentor observation, and citation-backed evidence, pre through post.

SOPACT · LEARNING INTELLIGENCE

Stop measuring training activity. Start proving the learner changed.

One persistent learner ID at enrollment. Every instrument inherits it — pre, mid, post. Level 3 and 4 become default outputs, and the funder report is a four-minute query, not a six-week project.

Map your architecture →

sopact.com

Training Assessment | Training Effectiveness | Training Evaluation | Kirkpatrick Model