PROGRAMME BOOK SERIES · No. 66

BEYOND THE CLOSURE REPORT

Learning From What Actually Happened: A Working Method for Post-Programme Evaluation

A step-by-step method for turning a closed programme into decisions the next one will actually use.

Giovanni Leonardi

May 2006

Contents

Learning From What Actually Happened: A Working Method for Post-Programme Evaluation

Programme · No. 66

Originally written May 2006 | Re-edited and re-rendered for the WhereBeyond Edition,

> *"A lesson is not learned when it is written down; it is learned when the next decision is made differently because of it."*

1. Executive Summary

Most programmes are evaluated once, badly, in the fortnight after they close, by the people least able to be honest about them and least available to do the work. The result is a lessons-learned log that is filed, admired briefly, and never consulted — and an organisation that pays full price for the same mistake three programmes running. This paper sets out a working method for doing it properly: a disciplined, seven-phase evaluation that begins before closure, reconstructs what the programme actually promised, establishes what it actually delivered, explains the gap through evidence rather than blame, and — the step almost everyone omits — embeds the findings in the governance of future programmes so that the next commissioning decision is genuinely different.

The method is deliberately modest in its machinery and demanding in its discipline. It requires an independent evaluation lead, a sponsor with enough seniority to make the findings land, a reconstructed baseline drawn from the original business case and benefits profiles, a structured pass over evidence and testimony, an honest benefits realisation assessment that names dis-benefits and emergent gains as well as planned ones, and a small number of hard gates that stop the evaluation degrading into a comfortable narrative. What follows is the method, phase by phase, with the roles, inputs, outputs, decision points and artefacts a practitioner needs to run it in a real organisation.

2. Why Most Post-Programme Reviews Teach Nothing

The post-implementation review, as it is usually practised, fails for reasons that are structural rather than personal. It is commissioned too late, so the people who could explain the early decisions have already moved on. It is run by the programme team, who cannot be expected to indict their own choices. It is scoped as an event — a workshop, a template, an afternoon — rather than as a piece of analysis. And it stops at the writing of the lessons, treating the document as the deliverable when the document is merely the evidence that the real deliverable, a change in future behaviour, might one day occur.

The consequence is a particular and expensive form of organisational amnesia. The benefits case that was approved is never revisited against what materialised, so the optimism baked into every business case is never disciplined by feedback. The same integration is under-estimated, the same stakeholder is forgotten, the same dependency bites, programme after programme, because nothing in the machinery forces the memory of the last one into the planning of the next. An evaluation that does not change a subsequent decision has not been performed; it has been performed at.

> *A lesson is not learned when it is written down. It is learned when the next decision is made differently because of it. Every step of this method exists to close the distance between those two moments.*

3. What This Method Produces

Run properly, the method yields five artefacts, each with an owner and a defined use. The Evaluation Mandate authorises the work and fixes its scope and independence. The Reconstructed Baseline states what the programme actually committed to, recovered from the original business case, blueprint and benefits profiles rather than from memory. The Benefits Realisation Ledger records, benefit by benefit, what was planned, what materialised, what did not, and what emerged unplanned — including dis-benefits. The Evaluation Report explains the variance through evidence and traces it to causes. The Decision Recommendations convert those causes into specific changes to how future programmes are commissioned, governed and staffed — each one addressed to a named part of the governance system.

The first four are common, if rarely done well. The fifth is the one that distinguishes an evaluation from a post-mortem, and the method is engineered around protecting it.

4. Roles and Responsibilities

Four roles carry the method; a fifth is consulted throughout.

- Evaluation Sponsor. Ideally the portfolio director or a Senior Responsible Owner unconnected to the programme under review. Owns the mandate, holds the gates, and — crucially — owns the onward decisions. Without a sponsor senior enough to act on the findings, the evaluation is an essay.
- Evaluation Lead. Independent of the programme's delivery line. Runs the analysis, owns the evidence standard, and is accountable for the report saying what the evidence supports rather than what is comfortable.
- Benefit Owners and Business Change Managers. The people into whose operation the benefits were supposed to land. They are the primary source for what actually

changed on the ground, and the primary check on any claim that a benefit was realised.

• Programme Office. Supplies and curates the documentary record — baselines, change logs, financial actuals, risk history — and maintains the lessons and evidence registers.

• Former Programme Leadership (consulted). The outgoing SRO and programme manager are interviewed as witnesses, not examiners. Their account is evidence to be weighed, not the verdict.

The separation of the Evaluation Lead from the delivery line is the single most important design choice. An evaluation written by the team that ran the programme will be, at best, a careful account of why everything that went wrong was unavoidable.

5. Phase One — Commission and Scope the Evaluation

Purpose. To authorise the evaluation, fix its independence, and define what it will and will not examine — before the programme dissolves.

Timing matters. This phase begins before the programme closes, not after. The trigger is the approach of closure, and the reason is brutally practical: once the programme is disbanded, the corporate memory disperses within weeks and the financial and delivery records begin to decay.

Inputs. The programme's closure plan; the original business case and its approved revisions; the portfolio's standing evaluation policy, if one exists.

Activities. The Sponsor commissions the Evaluation Lead and confirms their independence in writing. Together they set the scope: which objectives and benefits are in view, the period under examination, the depth expected, and — explicitly — what is out of scope, so the work does not sprawl into an inquest. They agree the evidence standard up front: what will count as proof that a benefit was or was not realised.

Output / artefact. The Evaluation Mandate: a two-page document stating purpose, scope, independence, evidence standard, timeline, and the onward decisions the findings are expected to inform.

Gate 1 — Mandate approved. The Sponsor confirms the scope is answerable with the time and access granted, and that genuine independence exists. If the Evaluation Lead is not independent of the delivery line, the gate does not pass.

6. Phase Two — Reconstruct the Baseline

Purpose. To establish what the programme actually committed to — not what people now remember it committing to.

This phase exists because memory is kind to the past. Ask a team six months after closure what the programme was for and you will get the version that the outcome flatters. The only defence is to recover the commitment from the contemporaneous record.

Inputs. The original and revised business cases; the blueprint or target operating model; the benefits map and individual benefit profiles; the approved cost and schedule baselines; the change-control log that moved them.

Activities. Reconstruct three things precisely. First, the intended outcomes — what the organisation was to be able to do that it could not before. Second, the benefits — quantified where they were quantified, with their original owners, measures and target dates. Third, the baseline of cost, time and scope, tracked through every approved change so that the programme is judged against what it was finally authorised to deliver, not the first number written down. Where the change log shows the baseline was quietly re-set to match reality, that itself is a finding.

Output / artefact. The Reconstructed Baseline: a single reference statement of committed outcomes, benefits and the final authorised cost/time/scope position.

Gate 2 — Baseline agreed. The former programme leadership are shown the reconstructed baseline and invited to challenge it on fact. Disagreement is recorded; the Evaluation Lead adjudicates on the documentary record.

7. Phase Three — Gather the Evidence

Purpose. To assemble, against the baseline, what actually happened — in numbers and in testimony.

Inputs. Financial actuals; delivery and milestone records; risk and issue history; operational performance data from the areas the benefits were to land in; and the people who were there.

Activities. Run two streams in parallel. The documentary stream collects hard data: actual cost against authorised cost, actual delivery against baselined schedule, and — most importantly — operational metrics from the receiving business that speak to whether benefits materialised. The testimony stream interviews a deliberately wide set of witnesses: not only the programme leadership but the benefit owners, the front-line staff who received the change, and at least one articulate sceptic. Every claim is logged with its source and its strength, so that the later analysis can distinguish a fact from an impression.

Output / artefact. The Evidence Log: each item tagged to a baseline element, its source, and an evidence grade (documented, corroborated, or asserted).

Gate 3 — Evidence sufficient. The Evaluation Lead confirms that the material is strong enough to support conclusions on the major benefits. Where it is not, the gap is itself reported rather than papered over with a confident sentence.

8. Phase Four — Analyse Variance and Causation

Purpose. To explain the gap between baseline and outcome, and to trace each significant variance to a cause that can actually be acted on.

This is where most reviews stop one step short. They establish that the programme ran twenty per cent over and six months late, and then they attribute it to a vague noun — "complexity", "stakeholder challenges" — that no future programme can do anything with. The discipline here is to keep asking why until the answer names something a commissioning body could have decided differently.

Activities. For each material variance, work forward through what happened and backward through why. Separate causes into three classes, because they demand different responses: decisions (choices that could have been made otherwise with the information then available), conditions (constraints that were real and known), and surprises (things genuinely not foreseeable). Only the first class yields transferable lessons; mislabelling a decision as a surprise is how organisations excuse themselves from learning. Test each proposed cause against the counter-factual: had this been decided differently, would the outcome plausibly have changed?

Output / artefact. A Causal Analysis section linking each significant variance to classified, evidenced causes.

Gate 4 — Causation defensible. The Sponsor tests whether the causes are supported by the evidence log and whether any comfortable narrative has crept in. Any cause graded only "asserted" is downgraded or dropped.

9. Phase Five — Assess Benefits Realisation

Purpose. To state honestly, benefit by benefit, what the programme actually delivered into the business — the question the whole exercise exists to answer.

Activities. Take each benefit from the reconstructed baseline and place it in one of five categories, with evidence: realised (delivered and demonstrable), partial (some value, short of target), unrealised (did not materialise), emergent (real value that was never planned), and dis-benefit (a negative consequence the change imposed on the operation). The last two categories are the ones ordinary reviews suppress, and they are where the truth usually lives. A programme that hit ninety per cent of its planned benefits while quietly imposing a dis-benefit on the front line has not been the success its closure report will claim.

Output / artefact. The Benefits Realisation Ledger:

Benefit	Planned measure	Actual	Category	Evidence grade
Reduced processing time	−30% by month 6	−19%	Partial	Corroborated
Retirement of legacy system	Full decommission	Not achieved	Unrealised	Documented
Staff redeployment saving	£480k p.a.	£310k p.a.	Partial	Documented
Improved data quality	Not planned	Material gain	Emergent	Corroborated
Increased front-line rework	Not planned	+2.5 hrs/week/team	Dis-benefit	Corroborated

Gate 5 — Ledger honest. The Benefit Owners confirm the ledger against their lived operational reality. A benefit no owner will personally stand behind is downgraded, regardless of what the programme claimed at closure.

10. Phase Six — Distil Transferable Lessons

Purpose. To convert the analysis into a small number of lessons general enough to apply to the next programme and specific enough to act on.

Activities. Resist volume. Twenty lessons are twenty lessons nobody will read; four or five that matter will change behaviour. For each candidate lesson, state the pattern observed, the evidence for it, the class of future programme it applies to, and — the test of a real lesson — the specific decision it should alter. A lesson that cannot name the decision it changes is an observation, and belongs in the report's narrative, not its recommendations.

Output / artefact. The Lessons Register entry set, each tied back to evidence and forward to a decision.

11. Phase Seven — Embed the Findings in Future Decisions

Purpose. To make the evaluation change something. This is the phase that justifies all the others, and the one almost universally skipped.

A lesson recorded in a register the next programme never opens has changed nothing. Embedding means routing each recommendation to the specific point in the governance system where the relevant decision is actually made, and giving it an owner there. A lesson about optimistic benefit estimates belongs in the business-case approval gate, as a mandatory challenge question. A lesson about under-scoped integration belongs in the assurance checklist the next programme must pass. A lesson about the wrong sponsor belongs in how the next SRO is appointed.

Activities. For each recommendation, name the governance mechanism it must modify, the owner of that mechanism, and the date by which the change takes effect. Convert lessons into standing questions and gate criteria rather than exhortations, because a criterion is enforced and an exhortation is admired. The Evaluation Sponsor personally hands the recommendations to those owners and secures their acceptance.

Output / artefact. The Decision Recommendations: each a triple of finding, governance mechanism to change, and accountable owner with a date.

Gate 6 — Findings embedded. The Sponsor confirms every recommendation has an accepting owner and a mechanism. Recommendations with neither are not closed as "noted"; they are escalated, because a finding with nowhere to land is a future repeat waiting to happen.

12. The Gates, in Summary

Gate	Held by	The question it asks
1 — Mandate approved	Sponsor	Is the scope answerable, and is the lead genuinely independent?
2 — Baseline agreed	Evaluation Lead	Does this reflect what was actually committed, on the record?
3 — Evidence sufficient	Evaluation Lead	Is the evidence strong enough to support conclusions?
4 — Causation defensible	Sponsor	Are the causes evidenced, or has a comfortable story crept in?
5 — Ledger honest	Benefit Owners	Will the people who received the change stand behind this?
6 — Findings embedded	Sponsor	Does every recommendation have a mechanism and an owner?

13. A Worked Illustration

Consider a shared-services consolidation, closed and declared a success, its closure report citing an on-budget delivery and the retirement of three legacy systems. Applied to it, the method tells a more useful story.

The reconstructed baseline recovers the original business case, which had promised a thirty per cent reduction in processing time and an annual saving of four hundred and eighty thousand pounds from redeployment. The evidence phase pulls the operational data the closure report never looked at: processing time down nineteen per cent, not thirty; redeployment saving running at three hundred and ten thousand, not four hundred and eighty. The causal analysis traces the shortfall not to "complexity" but to a specific, classifiable decision: the target operating model had assumed a single consolidated process, but a late concession — made to secure the sign-off of one reluctant division — preserved two local variants. That was a decision, made with the information then available, and it is the transferable lesson: concessions bought at the approval gate to buy a signature are paid for, with interest, in the benefits case.

The benefits ledger then surfaces what the celebration concealed — an emergent gain in data quality worth having, and a dis-benefit of roughly two and a half hours of new weekly rework per team, imposed by the very process variant that bought the signature. The recommendation writes itself, and it is not "manage stakeholders better": it is a standing question added to the business-case gate — which benefits does this concession put at risk, and who has priced them? — owned by the portfolio director. That is a closed programme turned into a sharper next decision. That is the point.

14. What Makes It Fail

The method fails in predictable ways, and naming them is part of running it. It fails when the Evaluation Lead is not truly independent, and the report drifts toward self-justification. It fails when it is commissioned after closure, and the evidence has already walked out of the door. It fails when it produces thirty lessons instead of five, guaranteeing none is acted on. Above all, it fails when Phase Seven is skipped — when the report is presented, thanked, and shelved, and no governance mechanism is actually changed. An evaluation that ends at the report is a cost. Only an evaluation that ends at a changed decision is an investment, and the whole of this method is built to carry the work across that final, most-neglected step.