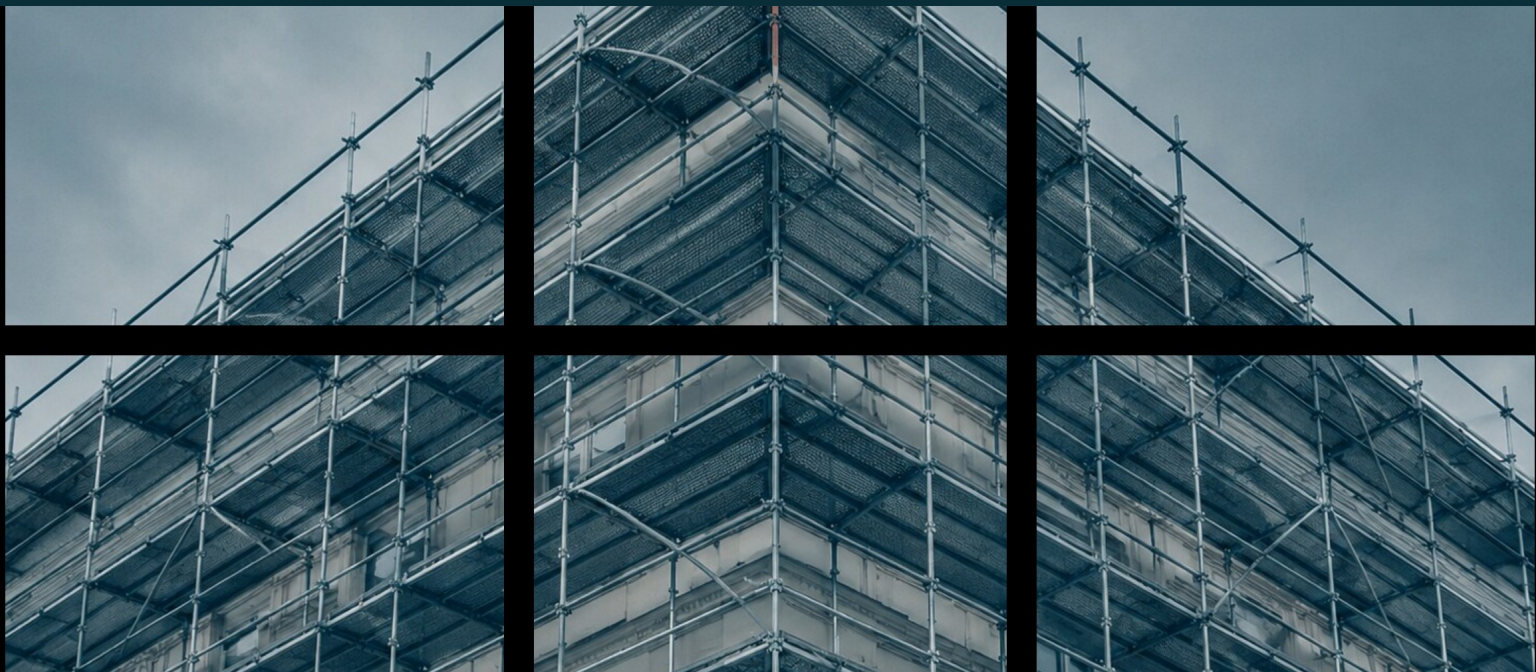


BEYOND THE BADGE

Why Maturity Models Measured the Wrong Things

How a decade of climbing capability ladders taught organisations to document change without getting better at it



Giovanni Leonardi

September 2006

Why Maturity Models Measured the Wrong Things

Transformation · No. 80

Originally written September 2006 | Re-edited and re-rendered for the WhereBeyond Edition, 2025

“Be suspicious of any measure you can pass without getting better.”

1. The audit passes, the programme slips

There is a particular quiet that settles over an organisation in the week of an appraisal. The binders have been assembled — process descriptions, tailoring guidelines, evidence of peer reviews held and defects logged and actions closed — and stacked in the room that has been booked for the assessors. People who have not spoken in months rehearse the same answers to the same questions. And when the verdict comes back and the organisation is confirmed at Level 3 — Defined — there is genuine, uncomplicated relief. A milestone has been reached. The letter goes to the board. The badge goes on the tender documents.

What is striking, if you have sat through more than one of these weeks, is how little relationship the celebration bears to the state of the actual work. In the same building, on the same afternoon, the programme that matters most to the organisation is sliding quietly from amber to red. Its steering committee is about to be told, for the third time, that the date has moved. Nobody in the appraisal room is thinking about it, because the appraisal was never really about that programme. It was about whether the organisation could describe how it works, and produce evidence that the description had been followed.

We have spent the better part of two decades building ladders to climb — the Capability Maturity Model and its integrated successor, and now the newer portfolio and programme maturity models arriving from the standards bodies — and we have climbed them with real diligence. The uncomfortable observation, which the appraisal week makes visible if you are willing to see it, is this: an organisation can climb the ladder while getting no better, and sometimes while getting worse, at the thing the ladder was supposed to measure. The models measured the wrong thing. They measured conformance to a defined process. They did not measure the capability to deliver an outcome. And those two things are not the same — they can, under enough pressure, become opposites.

2. What the ladder actually measured

It is worth being precise about the object of measurement, because the whole argument turns on it. A maturity appraisal asks, in effect, three questions of a process. Is it defined — written down, standardised, owned? Is it followed — is there evidence, in the form of artefacts, that people did what the definition says? And is it institutionalised — trained, resourced, audited, improved? These are not foolish questions. An organisation that cannot answer them is usually an organisation that delivers by heroics, where success depends on which individuals happen to be in the room, and where nothing learned on one project survives to the next.

But notice what every one of those three questions has in common. Each is answerable from documentation. Each rewards the visible and the recordable. The maturity of a process, as the model scores it, is a property of its paperwork — the existence of the procedure, the trail of evidence that the procedure was enacted, the minutes of the reviews that governed it. What the model cannot see, because it has no instrument for it, is whether any of this produced a better decision, a better product, or a better outcome for the organisation that paid for it.

A maturity model measures whether you can describe how you work and prove you followed the description. It is silent on whether the description was any good, and silent on whether following it made anything better.

This is the gap the whole edifice rests over, and it is wider than practitioners like to admit. Two organisations can hold the same rating and be nothing alike. One has reached Defined because its processes genuinely encode hard-won judgement, and its people use them because they help. The other has reached Defined because a process improvement group, under deadline, wrote three hundred pages of procedure that the delivery teams tolerate, evidence for the assessors, and quietly route around when real work has to be done. The model scores them identically. It has no column for the difference, and the difference is the only thing that matters.

3. The number that did not move

Let me put a concrete shape on this, because the abstraction is too comfortable. Consider — the details are composite, but the shape is one I have watched more than once — a large insurer that set out to reach Level 3 across its IT-enabled change function. It was a serious effort, properly resourced: a process improvement group of around fourteen people, seconded for the better part of fifteen months; a library that grew to something over three hundred defined procedures and templates; a cycle of internal appraisals leading to the formal one. The badge was achieved. It was announced. It was, by the terms it set itself, a complete success.

Now set beside that the measure the organisation actually cared about before the programme began: the proportion of its major change initiatives that landed the benefits their business cases had promised. Before the maturity effort, that figure sat at roughly two in five — three in five initiatives quietly failing to deliver what had been signed off. Eighteen months and a Level 3 badge later, it had not meaningfully moved. The organisation had become measurably better at documenting how it ran change, and no better at all at running it.

The tell was in what happened on the one programme that genuinely mattered that year — a core policy-administration replacement, the kind of undertaking on which the firm's next decade quietly depended. Faced with a delivery window that the defined process could not fit inside, the programme did exactly what a rational team does: it invoked the tailoring guidelines and stripped the standard process back to a fraction of itself. This was not a violation. The model permits tailoring; the paperwork was immaculate. But pause on what it means. On the single most important initiative in the portfolio — precisely the case the discipline exists for — the mature, defined process was formally set aside because following it would have made things worse. The process was fit to be shown and unfit to be used, and everyone involved knew the difference.

“The badge said Defined. The real work happened in the margins the badge could not see.”

That is the mechanism, and it is worth naming plainly because it is not an accident of one insurer. When you make conformance to a process the measure, you make conformance to a process the goal. People are not stupid; they optimise for what is inspected. The energy of the organisation flows toward producing evidence — the reviews that leave a trail, the artefacts that appraise well — and away from the harder, unrewarded work of judgement, of deciding what genuinely needed doing on this problem. A measure that can be satisfied without the underlying thing improving will, given time and incentive, be satisfied without the underlying thing improving. We did not need to wait for the maturity movement to learn this; economists had said as much a generation earlier. We simply declined to apply it to ourselves.

4. The case for the models — and where it breaks

I want to state the opposing case at its strongest, because it is a good case and the argument is worthless without it. Before the maturity models, much of this profession ran on chaos and charisma. There was no shared vocabulary for capability, no way for a buyer to distinguish a supplier who could deliver reliably from one who merely presented well, no ladder for an improving organisation to orient itself against. The Capability Maturity Model professionalised a field that badly needed it. It made the invisible discussable. It gave a genuinely immature organisation — one where every project reinvented its own approach and nothing was ever learned twice — a sequence of concrete, sensible next steps. And that first stretch of the climb, from the pure firefighting of Level 1 to the basic managed discipline of Level 2, delivered real and defensible value. Repeatability is not nothing. For an organisation drowning in its own inconsistency, it is a great deal.

All of that is true, and I would not want it undone. But notice where the value sits. It sits at the bottom of the ladder, where the problem is the absence of any process at all, and where defining one is genuinely the thing most worth doing. The trouble is that the model presents itself as a single monotonic good — that more maturity is more better, all the way up — and this is where it misleads. The value of definition does not increase smoothly as you climb. It peaks early and then inverts. Near the top of the ladder, the marginal effort goes not into capability but into the apparatus of demonstrating capability: the measurement of the measurement, the optimisation of the optimising. An organisation can spend its scarcest resource — the attention of its best people — on ascending from a high level to a higher one, and buy almost nothing an outcome would recognise.

The error, to be clear, was never in having a model. It was in mistaking the map for the territory. Process is defined is a fact about the map. This organisation can deliver a hard change under real uncertainty is a fact about the territory. The maturity movement's foundational move was to treat the first as a reliable proxy for the second — and then to build a whole industry of appraisal on the assumption that the proxy would hold. It holds at the bottom of the ladder. It comes apart at the top, exactly where the badges are most prized and most expensive.

5. Measuring the thing, not the ritual

If the diagnosis is right, the response is not another, better-calibrated maturity model — a more elaborate map is still a map. The response is to move the measurement onto the territory, and to be willing to measure things that are harder to score precisely but actually correspond to capability. It is better to be roughly right about the outcome than precisely right about the paperwork.

Consider the contrast directly:

What the maturity appraisal counts	What actually predicts delivery
The process is documented and owned	The process is used when the deadline is real, not tailored away
Evidence exists that reviews were held	The reviews changed a decision that would otherwise have been wrong
The organisation has reached a defined level	The organisation's initiatives realise the benefits they promised
Capability is institutionalised on paper	Capability survives the departure of the three people who really hold it

None of the items in that right-hand column appraises cleanly. You cannot certify them in a week with a room of binders. They require watching an organisation under load — asking not is there a process but what does this place do when the plan meets reality and something has to give. That is a slower, more honest, and less flattering form of assessment, which is precisely why the maturity models displaced it: a ladder with a number on it is easier to put in a board pack than a judgement about whether the organisation can actually cope.

So the practical discipline is a change of question. Not what level are we, but what have we become better at that a customer or a board would notice. Track the benefits initiatives actually realise against the benefits they promised, and watch the trend. Notice whether your defined process is genuinely reached for under pressure or quietly abandoned the moment a date is at risk — the second is the more revealing signal, and it is free to observe. Ask, of any capability you claim to have institutionalised, whether it would survive the resignation of the few people who really carry it. These are cruder instruments than a five-level scale. They have the singular advantage of measuring the thing itself rather than the ritual that surrounds it.

6. The test worth keeping

There is one habit from all this worth carrying forward, and it is close to the opposite of what the appraisal week trains. Be suspicious of any measure you can pass without getting better. That suspicion is the whole of it. The maturity models failed not because measurement is wrong — a profession that refuses to measure itself stays a craft forever — but because they measured the residue of capability, the paperwork it leaves behind, and mistook that residue for the thing.

The organisations that got the most from the models, in my experience, were the ones that used the climb as a scaffold and then had the confidence to let it go — that took the discipline of Level 2, the shared language and the basic repeatability, and declined to

spend their best years chasing the badges above it. They kept asking the harder question. Meanwhile the room with the binders in it goes quiet, the letter goes to the board, and down the corridor the programme that will actually decide the next decade slips one more time, unmeasured, because no ladder we built ever had a rung for it.