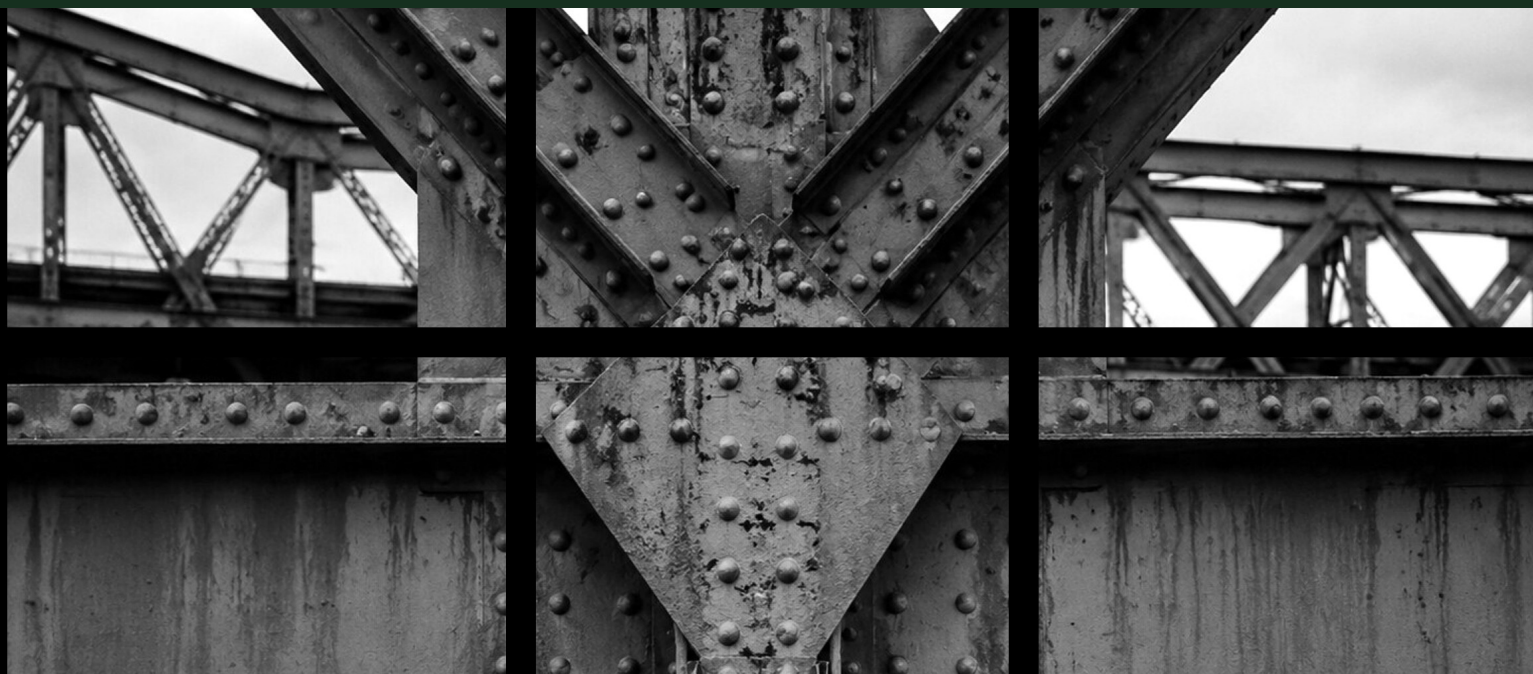


BEYOND THE GREEN DASHBOARD

When the Measure Becomes the Target: Why Programme Metrics Corrupt Under Pressure

On Goodhart's Law, the reporting reflex, and what a decade of green status reports teaches us about the numbers we choose to trust.



Giovanni Leonardi

May 2007

When the Measure Becomes the Target: Why Programme Metrics Corrupt Under Pressure

Programme · No. 80

Originally written May 2007 | Re-edited and re-rendered for the WhereBeyond Edition, 2025

“A team punished for red will show you green; a team thanked for surfacing red early will show you the truth while there is still time to act on it.”

1. The report was green

Every indicator on the programme dashboard was green. Milestone attainment stood at ninety-two per cent, earned value was within tolerance, and the steering committee — which met for ninety minutes once a month, and read the pack in the ten minutes before — noted good progress and moved on. Four months later the programme missed its go-live by two quarters, and the review asked the question every review asks: how did no one see it coming?

They saw it. They had been looking at it for a year. What they were looking at had simply stopped meaning what they believed it meant.

This is not a story about a careless board, a dishonest team, or a badly built dashboard. It is a story about what happens to almost any number the moment an organisation decides to steer by it — and because it happens so reliably, across so many programmes and so many sectors, it deserves to be treated not as bad luck but as something close to a law.

2. What Goodhart actually said

In 1975 the economist Charles Goodhart, writing about monetary policy, observed that any statistical regularity a government relies upon for control will tend to collapse once the pressure to control it is applied. Two decades later the anthropologist Marilyn Strathern compressed the idea into the sentence most of us now quote: when a measure becomes a target, it ceases to be a good measure.

The programme world has absorbed the slogan and missed the mechanism. We treat Goodhart's Law as a witticism about bureaucrats, something that happens to other people's numbers. It is in fact the central operating risk of every reporting regime we run. The moment a metric is promoted from description — this is roughly where we are — to target — this is where you must be by Friday — the people whose performance it measures acquire both the motive and the means to make the number move without moving the reality beneath it. Not through fraud, in the overwhelming majority of cases. Through the thousand small, defensible, locally rational adjustments that a number invites the instant a career is attached to it.

A metric is a promise that the map still resembles the territory. A target is an incentive to redraw the map.

3. The reclassification reflex

Consider the most respectable metric in programme management: on-time milestone completion. It is objective, it is auditable, and it is almost perfectly gameable.

On one large integration programme I watched the on-time figure hold steady above ninety per cent for the better part of a year while the delivery date quietly walked backwards. There was no falsification. There was re-baselining. A milestone due in March, visibly slipping, was formally re-planned to June before the reporting cut-off — so it was never recorded as late; it was recorded as re-scoped, by agreement, through proper change control. In the first quarter of the programme there had been three such re-baselines. By the third quarter there were more than forty. The on-time percentage, computed against the current baseline, stayed green throughout, because the baseline was the thing that was moving. The number was not lying. It had simply been taught to keep up.

The same reflex has a dozen dialects. In test phases measured by defects closed, defects are reclassified as change requests or as "known issues, deferred to operations," and the burn-down chart descends beautifully into a live service that lists on day one. Where the metric is test cases executed, the easy cases are run first and in volume, so the count climbs while the coverage that matters waits until there is no time left. Where cost variance is the watched figure, spend is capitalised, deferred, or parked against a contingency line that no one reports on. Each move is individually justifiable. Each would survive an audit. Together they hollow out the very signal the board is relying on to sleep at night.

4. Why it is structural, not a failure of character

It is tempting to read all this as a morality tale — weak teams, poor discipline, a culture that needs tightening. That reading is comfortable and wrong, and it is wrong in a way that guarantees the problem recurs.

The corruption is structural, and it has three sources. The first is distance. The further a number travels from the work it describes, the more it is trusted and the less it can be interrogated. By the time a figure reaches the steering committee it has been aggregated through four layers, and each layer has rounded, smoothed, and framed it in perfectly good faith. The board sees a clean green cell precisely because it can no longer see the mess the cell was distilled from.

The second is pressure, which is Goodhart's own word. A measure that no one cares about stays honest indefinitely. It is the act of caring — of tying funding, reputation, or the next tranche of a bonus to the number — that sets the corruption in motion. We tend to apply the most pressure to exactly the metrics we most need to remain truthful.

The third, in this particular era, is a reporting culture that has learned to mistake documentation for control. Since Sarbanes-Oxley, and with Basel II now working its way through every financial-services change portfolio, we have built programmes that

generate more status, more attestation, and more dashboard than at any point I can remember. The paradox is that the volume of reporting has grown while its information content has thinned. A board drowning in green cells is not better informed than one with a single honest amber; it is merely more comprehensively reassured.

*“We have never had more numbers about our programmes,
and never been easier to fool with them.”*

5. "You cannot manage what you cannot measure"

The strongest objection to everything above is also the oldest, and it deserves a proper hearing rather than a straw-man dismissal. It runs: without hard targets there is no accountability. Measurement is the discipline that drags a programme out of the fog of "good progress" and "nearly there." Remove the target and you do not get honest description — you get drift, optimism, and a delivery date defended by feeling rather than fact. The balanced scorecard, earned value, the whole apparatus of quantified control exists because the alternative — managing by anecdote and confidence — failed more often and more expensively than any gamed metric ever has.

This is correct, and it is why the answer is not to abandon measurement. It is to be honest about what measurement is for. A number used to understand a programme and a number used to govern the people running it are doing two different jobs, and the tragedy is that we habitually use the same number for both. The instant the figure that helps you see the work becomes the figure that determines who is rewarded, you have converted your best instrument into your least reliable one. The discipline the objection rightly demands is real; it simply cannot be delegated to a single number under maximum pressure and expected to survive.

6. Measures that keep their meaning

If Goodhart's Law cannot be repealed, it can be managed, and the practitioners who manage it best tend to do a few unfashionable things.

- They measure in baskets, not single figures. A milestone metric paired with an independent read of downstream defect rates is far harder to game than either alone, because the moves that flatter one tend to expose the other. Gaming a single number is easy; gaming a well-chosen pair in opposite directions is real work, and usually not worth it.
- They keep the measurer close to the work. The number that has passed through four layers should be checked, occasionally and unpredictably, by someone who walks down to where the work actually happens and looks. A single afternoon spent watching a "ninety per cent complete" module being tested will tell you more than a year of the cell that says ninety per cent.
- They separate the measure from the reward wherever they can. The metric used to learn should not be the metric used to pay. When the two must coincide, they treat every impressive movement in the number with the suspicion it has earned.

- They treat variance as information, not sin. A team punished for red will show you green; a team thanked for surfacing red early will show you the truth while there is still time to act on it. The organisations that see trouble coming are, without exception, the ones that made it safe to report.
- They retire metrics on a schedule. A measure has a useful life. Once everyone has learned to optimise for it, it has quietly become a target in Goodhart's sense, and should be rotated out before it does more harm than good.

None of this is a methodology, and it does not come shrink-wrapped. It is a temperament — a settled refusal to be comforted by a number simply because it is green.

7. The number and the nerve

The deficit that sinks programmes is rarely a deficit of measurement. Most struggling programmes are measured to within an inch of their lives. What they lack is the nerve to look past the measure — to treat a wall of green not as reassurance but as a question, and to keep asking it when the pack says everything is fine.

Goodhart's Law is not a reason to stop measuring. It is a permanent reminder that our instruments are only ever as honest as we make it safe for them to be. The programmes that realise their benefits are not the ones with the best dashboards. They are the ones led by people who never quite believed the dashboard, and had the temperament to go and check.