

BEYOND COLOUR

RAG Status Measures Reporting Discipline — Not Programme Health

Why milestone counts and traffic lights conceal the decisions, dependencies and benefits that matter



RAG Status Measures Reporting Discipline – Not Programme Health

Programme · No. 112

Originally written April 2008 | Re-edited and re-rendered for the WhereBeyond Edition, 2025

“A metric matters only when it changes a decision, an action or the timing of either.”

1. The board pack that said green

The programme had twenty-two workstreams, a £96 million budget and a board pack that opened with an encouraging fact: 82 per cent of milestones had been achieved on time. Seventeen workstreams were green, four amber and one red. The charts were consistent, the tolerances documented and the trend stable.

The implementation date still moved by eleven weeks.

Nothing in the pack was false. That was the troubling part. Milestones had been counted as achieved when design documents were approved, even though unresolved interfaces remained. Workstreams were green because each could meet its own dates if shared specialists arrived when requested. The risk register recorded the resource pressure, but no measure showed that six plans depended on the same four people during the same fortnight. The programme had measured progress faithfully and health badly.

This pattern recurs because the measures most easily standardised are not necessarily the measures most useful to decision-makers. Red, amber and green status, milestone achievement and budget variance create a common language. They do not, by themselves, reveal whether the programme's promises are becoming more or less credible.

The textbooks teach us how to measure control. They say much less about measuring the quality of the programme's judgement.

2. RAG is a conclusion disguised as evidence

A RAG status looks like data, but it is usually a compressed opinion. Someone has taken several facts, applied thresholds and reduced the result to a colour. Compression is useful: senior leaders cannot inspect every schedule line and cost transaction. But compression also removes the reasoning that matters.

Two projects can both be amber for entirely different reasons. One has a known four-week delay with a funded recovery plan and a decision due tomorrow. The other is on its reported date but depends on an untested supplier estimate, an unavailable operational team and three overdue decisions. The first may be under control. The second may be far more dangerous.

Colour does not tell us:

- how reliable the forecast has been;

- which assumption carries the date;
- whether the variance is recoverable;
- how long an essential decision has been waiting;
- whether a local recovery transfers harm elsewhere;
- whether the intended benefit still justifies the remaining cost.

The absence of these questions turns status into ceremony. Project managers learn the tolerance rules, preserve green as long as the rules allow and declare amber when recovery is already expensive. The PMO aggregates the colours accurately. The board receives a precise account of yesterday's reporting convention.

I have seen programmes spend days debating whether a workstream is amber or red while leaving the underlying decision untouched. The colour attracts attention because it is visible; the decision escapes because nobody has measured its age, owner or consequence.

3. Milestones reward motion, not necessarily progress

Milestones appear more objective. A date was met or it was not. Yet milestone measurement is only as meaningful as the event being counted.

The weak programme measures the production of artefacts: requirements signed off, design completed, test plan approved, training materials issued. These events may matter, but each can be achieved while the programme's real uncertainty remains unchanged. Approval can mean that a document passed through governance, not that the organisation is ready to act on it.

The stronger question is not simply whether a milestone occurred. It is whether crossing it reduced uncertainty, released dependent work or made an outcome more credible.

Consider the composite programme again. Twelve design milestones are reported complete. A short review finds:

- five designs still contain assumptions awaiting operational confirmation;
- three share an interface specification that has not been tested;
- two require supplier changes not yet priced;
- four depend on data conversion rules owned by nobody.

The 82 per cent achievement figure is arithmetically correct. It is operationally misleading because completion has been separated from consequence.

This does not mean abandoning milestones. It means distinguishing event completion from readiness evidence. A design milestone should not be treated as complete merely because a paper is signed. The measure should also show whether its exit conditions are satisfied, what exceptions were accepted and which dependent work can now proceed without qualification.

4. Measure the promises, not just the plan

Every programme is a collection of promises: a date, a cost, a capability, an operational change and a benefit. Useful measures test the credibility of those promises.

A PMO should therefore ask five questions before adding any metric.

Question	What the measure should expose
Can we trust the forecast?	Accuracy of previous forecasts and movement in estimates
Can the programme make its decisions?	Age, ownership and consequence of unresolved decisions
Can the parts arrive together?	Dependency exposure and contention for shared resources
Can the organisation use what is delivered?	Readiness of people, process, data and operations
Is the remaining investment still justified?	Movement in benefit assumptions, cost and time to value

This shifts attention from reporting performance to programme behaviour.

Forecast reliability compares what teams predicted with what subsequently occurred. If a workstream repeatedly forecasts completion four weeks ahead and repeatedly misses, the issue is not one late milestone. It is that the forecasting mechanism cannot be trusted. A simple measure of date movement over the previous three reporting cycles often reveals more than the current colour.

Decision latency records the days between a decision becoming necessary and an authorised choice. It also records the cost or delay accumulating while the choice waits. A board may meet punctually while the programme remains slow because papers arrive without options, owners are absent or decisions are returned for more analysis.

Dependency exposure measures commitments that rely on another workstream, supplier or scarce specialist, particularly where the parties use different dates or assumptions. The useful unit is not the number of dependencies in a register. It is the value and timing of work exposed to unresolved dependencies.

Readiness confidence tests whether operational owners have accepted the conditions required for implementation. Training attendance alone is weak evidence. Signed procedures alone are weak evidence. Better evidence includes completed rehearsals, resolved exceptions, named support capacity and confirmed data quality.

Benefit credibility tracks whether the assumptions supporting the business case remain intact. A programme can deliver scope on time and still destroy value if volumes, adoption, operating cost or implementation delay change. Benefits should not remain untouched in the business case while every delivery assumption moves around them.

A metric matters only when it changes a decision, an action or the timing of either.

5. The strongest defence of simple metrics

There is a serious case for retaining RAG and milestone measures. Programmes already struggle to collect consistent information. Boards have limited attention. A small number of standard indicators enables comparison, escalation and accountability. Add too many measures and the PMO creates a new bureaucracy: more definitions, more reconciliation

and more arguments about data quality.

That defence is right. Complexity is not insight.

The answer is not a vast scorecard. It is a tighter distinction between navigation measures and diagnostic evidence. RAG, milestone achievement, cost variance and risk exposure can remain as navigation measures. They tell leaders where to look. But every important status must be supported by diagnostic evidence showing why the conclusion is credible and what decision follows.

A useful board pack may contain only seven or eight programme-level measures. The discipline lies in choosing measures that reveal a mechanism, not in filling every available space on the page.

A simple metric is valuable when it directs attention to the right question; it is dangerous when it pretends to be the answer.

The PMO should also retire measures. If nobody has changed a decision or action because of a metric in three reporting cycles, ask why it exists. It may be a compliance requirement, in which case label it as such. Otherwise it is probably inherited reporting rather than management information.

6. What measurement changes inside the PMO

Better metrics alter the office's work.

A reporting analyst stops spending most of the week correcting colours and starts testing forecast movement. A planner stops counting late milestones and examines whether accepted slippage consumes contingency or transfers delay across an interface. A risk coordinator stops reporting the number of open risks and shows which exposures have no funded response. The PMO manager stops presenting the pack as a finished product and uses it to frame the decisions the programme director must make.

This is not merely a technical improvement. It changes relationships. Project managers can no longer protect a local status by moving a consequence elsewhere. Sponsors can no longer ask for "more detail" without confronting the decision already available. Operational leaders become visible when readiness depends on their commitment rather than on the production of a plan.

It also requires courage from the PMO. Measures of forecast reliability may expose respected project managers. Decision-age measures may expose the board itself. Benefit credibility may challenge the original case for continuing. An office that is permitted to measure only delivery teams will produce a partial and politically convenient picture.

The honest test is symmetry: can the measurement system reveal failure in projects, suppliers, the PMO, the programme director and the board? If not, it is an instrument of supervision, not governance.

7. A better conversation than red or green

The most useful programme review I know begins with four sentences, not a colour:

- What changed?
- Why does it matter?
- Which promise is now less credible?
- Who must decide what, by when?

RAG status can follow. So can milestone trends, cost variance and risk counts. But those measures should support the conversation rather than define it.

The uncomfortable truth is that organisations often prefer simple status because it preserves the appearance of control. A green programme reassures. An amber one attracts questions. A red one demands intervention. The categories are administratively clear even when the programme is not.

Real control is less comfortable. It requires leaders to see uncertainty before it becomes failure, to compare options before the plan forces one upon them and to revisit benefits while there is still time to stop or reshape the work.

The PMO's contribution is not to produce more measures. It is to preserve the chain from evidence to interpretation, from interpretation to decision, and from decision to consequence. RAG and milestones remain useful servants within that chain. They become dangerous only when we mistake them for the whole of programme health.