



Whitepaper

Ten numbers

Measure leadership on the cross-functional work of getting new products out — not on functional performance alone.

Innovation series.

New products are the biggest lever a company has for growth, and the work of producing them runs horizontally across every function. Most company scorecards run vertically. This paper examines what that mismatch costs, drawing on a live lateralworks measurement engagement, and sets out a ten-metric framework on two clocks, including the one-page dashboard, that measures the leadership team on the thing that grows the company.

Prepared by

lateralworks
Performance measurement

Date

August 2026
Rev 1.0

Online

lateralworks.com
Innovation series

Table of contents

Ten numbers

01 The biggest lever, measured the wrong way	4
02 What a functional scorecard misses	8
03 Verdict on five proposed metrics	10
04 Ten numbers, on two clocks	14
05 The weekly six and the quarterly four	17
06 What is deliberately excluded, and why	20
07 The one-page dashboard	22
08 Making the numbers real	25
References	28

Core thesis. The scorecard is a strategy document whether or not it is written as one. If the strategy says new products and the scorecard says functional efficiency, the scorecard wins: people behave as they are measured. A company that wants new products out faster measures its leadership team together, cross-functionally, on ten numbers, six weekly and four quarterly, that make the product engine visible, and holds everything else in a register behind them.

Overview

Abstract

New products are the lifeblood of a company. Technology companies survive or die on their success, and every durable growth story in technology rests on a stream of them. Yet when management teams sit down to design a scorecard, the numbers that come back almost always run along functional lines: engineering efficiency, marketing spend, manufacturing cost, budget adherence. Each function is measured on its own vertical, and the horizontal work of moving a product from idea to revenue, work that crosses every one of those functions, belongs to nobody.

This paper takes up the question directly: should a company measure and reward functional performance, or cross-functional cooperation designed to get new products out faster? Six decades of measurement research warn against the first option, and the product development benchmarks endorse the second. Functional excellence still matters. But the headline scorecard — the dozen numbers a leadership team reviews together — has to sit on the horizontal flow, because that is where the growth is and because measurement placed anywhere else pulls behavior away from it.

The framework here comes from a live engagement. A mid-market hardware company, midway through a strategy shift toward connected products and recurring revenue, proposed five metrics to track company performance: cycle time reduction, new products released per year, performance to schedule against baseline, product quality, and time to break-even. All five are sound. All five are also execution measures, four of the five are lagging, and none would move if the strategy died. The replacement is ten metrics on two clocks: a weekly operating scorecard of six, reviewed by the CEO and product development leadership to change the date, and a quarterly business scorecard of four, reviewed by the portfolio committee to judge whether the strategy is working. Four of the five survive in the system, each sharpened; the fifth is replaced. The engagement is hardware, but the method, and most of the ten, transfers to any company whose growth depends on shipping new products. It is a worked example, not a sample.

Two design choices carry most of the value. The set is capped at ten, with everything else held in a register, each entry with a stated reason for exclusion and a rule for promotion. And the ten are published as a one-page dashboard the leadership team reviews as one body, in one meeting, rather than as functional reports reviewed one-to-one. The paper reproduces that dashboard, anonymized.

Key finding. At the client whose engagement produced this framework, the weekly six went live within eight weeks: three as measured values once the plan of record was reconciled, and the rest producing first real readings from a four-field decision log and the teams' daily boards, with no new systems and no new headcount. The constraint is almost never measurement capability. It is an agreed definition and a named owner for each number.

01

The lever

The biggest lever, measured the wrong way

Ask a leadership team what will grow the company and the answer is some version of new products: new platforms, new categories, new revenue streams layered onto the installed base. Ask the same team what its members are measured on and the answer changes shape. Engineering is measured on utilization and budget. Marketing is measured on demand generation. Operations is measured on cost and delivery. Quality is measured on escapes. Finance is measured on forecast discipline. Each measure is reasonable on its own terms, and together they describe a company that has organized its attention around everything except the thing it says will grow it.

This section makes the case that the mismatch is expensive, and that it is a measurement problem before it is an organization problem. The evidence comes from two directions: research on what measurement does to behavior, and benchmarking on what separates companies that are good at new products from companies that are not.

New products carry the growth

The strategic weight of new product development is not in dispute. BCG's 2024 innovation survey put 83 percent of companies ranking innovation among their top three priorities, while only about 3 percent were judged ready to convert that priority into results [1]. The Product Development and Management Association's global best-practice surveys, running for three decades, consistently find that the firms it classifies as best are set apart by how they combine and manage development practices, not by any single tool or spending level [2]. The pattern has deep historical roots. Tektronix grew from a \$1.2 million instrument maker in 1950 to one of the defining technology employers of the Pacific Northwest on the strength of a continuous stream of new instruments [3], and Hewlett-Packard institutionalized the same idea a generation later by making new-product economics a first-class management object [4].

A company's new product engine is also its most cross-functional machine. A product moves from customer evidence to concept to design to qualification to ramp to revenue, and every handoff crosses a functional boundary. The engine's speed is set by how well those crossings work, not by how well any single function performs inside its own walls. That is what makes the measurement question consequential: the work is horizontal, and most measurement is vertical.

Measurement is not a mirror

The research on performance measurement has been saying the same thing since the 1950s: measures do not just report behavior, they produce it. Ridgway's 1956 survey of measurement systems catalogued how single, composite and multiple indicators each distort effort toward whatever is counted [5]. Kerr's 1975 classic "On the folly of rewarding A, while hoping for B" documented organizations hoping for teamwork while rewarding individual results, and the paper remains uncomfortable reading for any company that hopes for cross-functional cooperation while paying for functional performance [6]. Kaplan and Norton opened their 1992 balanced scorecard article with the sentence this paper borrows as its epigraph: what you measure is what you get [7]. Hauser and Katz formalized the mechanism: teams optimize the metric they are given, including all of its defects [8].

lateralworks reached the same conclusion from field observation before connecting it to the literature. The published research on integrated core teams notes that goals and measurement systems favoring vertical communication over horizontal cooperation are one of the reliable ways companies hollow out their own cross-functional teams: members identify with the function that controls their raise and their next role, and the team becomes a meeting rather than a unit [9]. People behave as they are measured, and they are almost always measured by their function.

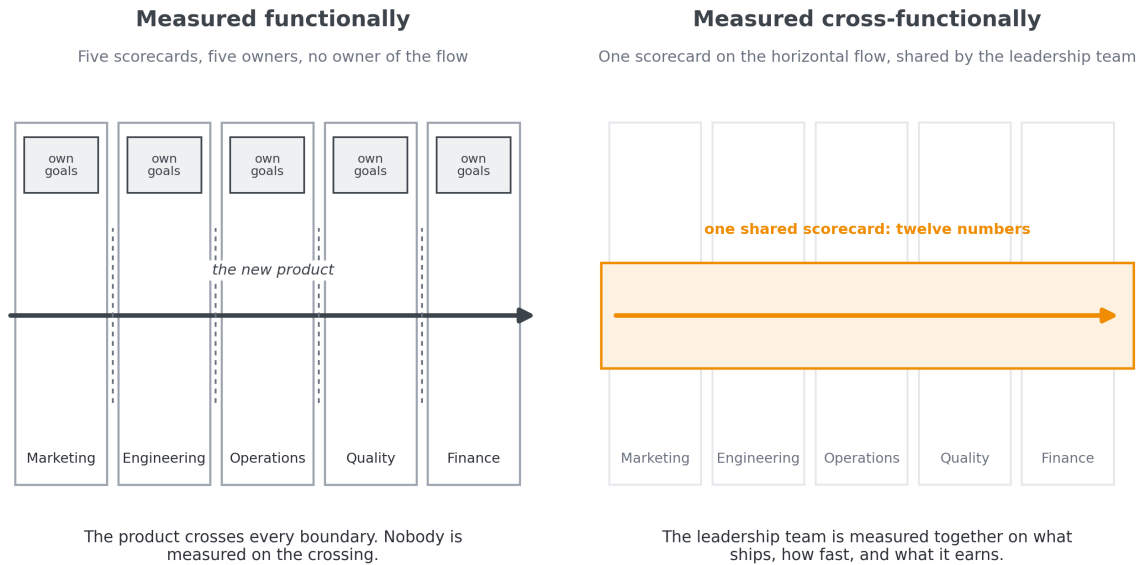


Figure 1. The same five functions, measured two ways. A functional scorecard puts a number on every vertical and none on the flow that generates the growth. A cross-functional scorecard measures the leadership team, together, on the flow.

What the benchmarks show

The product development literature closes the loop. In the APQC benchmarking study by Cooper, Edgett and Kleinschmidt, genuinely cross-functional project teams (technical, marketing, sales and operations people on one team, with one leader) were present in 79.3 percent of the best-performing businesses and 7.7 percent of the worst [10, 11]. That is a correlation, not a proof of cause; the case in this paper rests on the mechanism, and the benchmark says the mechanism is worth betting on. Ancona and Caldwell add the caution that composition alone does nothing: cross-functional membership pays off only when the team has integrating processes and shared goals to go with it [12]. Katzenbach and Smith drew the sharper line: a real team is defined by mutual accountability to common performance goals, and a group of executives each accountable only for a functional result is not a team but a working group [13].

The cautionary tales sit on the other side of the ledger. 3M's new product vitality index, the share of revenue from products launched in the past five years, had historically run at a third or better; after several years of efficiency-first management it slid toward a quarter, and the recovery under new leadership was tracked with the same index [14, 15]. McKinsey's study of product development metrics found that 85 percent of companies track adherence to budget, and that budget adherence was the only metric in the study with a significant negative correlation to both short-term profit growth and long-term stability [16]. Companies measure what is easy to count inside a function, and the measure quietly works against the outcome.

The question, stated plainly

So: should a company measure and reward functional performance, or cross-functional cooperation designed to get new products out faster? Functions still need operating measures: a factory has yield, a quality lab has turnaround time, and nothing here removes them. The question is what goes on the headline scorecard, the dozen numbers the CEO and the leadership team review together and tie consequences to. The answer this paper defends is that the headline set belongs to the horizontal flow. The rest of the paper builds that set: what a functional scorecard misses, the verdict on the metrics a management team usually proposes, the ten that replace them, the discipline of what stays off the wall, and the one-page dashboard that puts the ten in front of the leadership team as one shared instrument.

02

The gap

**What a functional
scorecard misses**

The client behind this framework is a mid-market hardware company; its identity is withheld and the details below are lightly generalized [17]. The shape that matters: a profitable, category-leading instruments maker, a product line being substantially replaced over three years, and roughly ten product development teams running in parallel. Its strategy sets out three capability pillars: own the customer intelligence, own the system architecture, and execute with excellence. On top of the pillars sit two platforms (a modernized core product platform and a connected digital platform) and a recurring-revenue ambition with a five-year ramp. The management team, to its credit, asked for a scorecard. It proposed five metrics: cycle time reduction, new products released per year, performance to schedule against baseline, product quality, and time to break-even.

All five are legitimate. All five also measure the same thing: whether the development engine converts ideas into shipped hardware. That was the right scorecard for the company it has been. It is roughly half the scorecard for the company its strategy says it is becoming.

A development scorecard presented as a company scorecard

Read the five proposed metrics against the strategy and the asymmetry is immediate. Every one of them lives inside “execute with excellence.” None of them tells the management team whether the company is winning on customer intelligence, the capability its strategy names as hardest for competitors to copy. None measures whether it is coming to own its system architecture or whether that ownership still sits with outside design partners. And with the partial exception of break-even time, none of them would move at all if the entire digital platform were cancelled tomorrow.

Strategic commitment	Covered by the proposed five?	What is missing
Own the customer intelligence	No coverage.	Nothing measures whether customer understanding is deepening or whether product decisions are grounded in evidence.
Own the system	No coverage.	No measure of IP ownership versus partner dependency, platform reuse, or architecture leverage.
Execute with excellence	Full coverage — this is the whole set.	Mostly lagging. No leading indicators of schedule risk, decision speed, or team capacity.
The connected platform	No coverage.	Recurring revenue, subscription retention, adoption and installed connectivity are all absent.

Figure 2. The five proposed metrics read against the company's own strategic commitments. Everything sits in one pillar.

Steering by the wake

The second structural problem is timing. Cycle time, products per year and time to break-even are all retrospective: they report the verdict after the money is spent. Performance to schedule is retrospective too unless it is paired with a forward view. A management team that receives only lagging indicators is being asked to steer by looking at the wake. lateralworks has published the counter-practice for years: track the gap between the bottom-up plan and the committed target, weekly, so slip is visible as it forms rather than at the next gate [18]. On this program, decision speed was the single named environmental weakness in the assessment record; a scorecard made of lagging measures would have reinforced it, because nothing on the page creates a reason to decide today.

There is also a subtler cost. When the headline set is functional and retrospective, review meetings become serial status readouts: each executive reports a number the others cannot act on, and the room disperses. The alternative developed in the rest of this paper gives the room a shared set of forward-looking numbers, so the meeting has decisions in it rather than narration.

03

Assessment

Verdict on five proposed metrics

Rejecting the five proposed metrics outright would have been both wrong and politically foolish: they are sensible measures, and the team that proposed them was reasoning from a genuine picture of the business. The right treatment is a verdict on each: keep four, replace one, and redefine three of the four that stay. This section gives the verdicts and the reasoning, because the redefinitions carry most of the value and they generalize to any company designing a product-engine scorecard.

#	Proposed metric	Verdict	What has to change
1	Cycle time reduction	Recast forward	The backward-looking duration becomes the forward-looking weekly gap to target; the front end it was meant to expose is guarded by customer evidence coverage instead.
2	New products released per year	Replace	Substitute the vitality index. Counting products penalizes platform work and is trivially gamed by splitting SKUs.
3	Performance to schedule vs. baseline	Keep, recast	Gap to target becomes the weekly headline; forecast accuracy is read as its six-week trend rather than published as a second metric.
4	Product quality	Keep, split	Field and ecosystem quality separate, reviewed at the gates and the monthly quality review. Ecosystem telemetry still locks at Release 1.
5	Time to break-even	Keep, at the gate	Set with cost of delay at the funding gate and owned by the team. Portfolio yield carries the money question on the quarterly page.

Figure 3. Verdict on the five proposed metrics.

Cycle time: the clock starts at the idea

Cycle time is only honest if the clock starts when the idea is first discussed rather than when a funded team forms. The lateralworks FTTM research is unambiguous on this point: much of a program's eventual overrun is lost in the fuzzy front end, the gestation period between idea and staffed team, which in unmanaged organizations consumes half to all of the eventual development cycle [19, 20]. Smith and Reinertsen made the same observation two decades earlier: the front end is where time is cheapest to save and where nobody is watching [21]. Measure from the funding gate and the largest single source of delay in the company becomes invisible. The two-clock system answers with position rather than duration: customer evidence coverage counts the front-end work before the commit, and the weekly gap to target watches the development clock while there is still time to act. Duration at closure stays in the register; it is history by the time it prints.

Products per year: replace it with the vitality index

Products released per year is the weakest of the five and the only one worth removing. It counts activity rather than value, it is gamed by splitting one product into three SKUs, and it actively penalizes the work the strategy depends on: platform and infrastructure programs produce no countable products while gating every product that ships. The vitality index, the share of revenue from products launched in the last three years, captures the same management intent weighted by value. It has a long pedigree at 3M, where it has served for decades as the company-level number for innovation health [14], and it was already in this client's R&D; strategy with a target near twenty percent. The vitality index can be gamed too, by refreshing a SKU and calling it new, which is why the definition settled in section 08 counts distinct product family launches only.

Schedule: promote the leading half

Performance to schedule against baseline rewards conservative baselines and punishes honest replanning, and it assumes a single agreed baseline exists. At this client it did not: two internal planning views disagreed on the scope of the later releases, and no rolled-up critical path existed across the portfolio [17]. The recast is to promote gap to target — the bottom-up critical-path finish date against the committed release date, refreshed weekly — to the weekly headline, read forecast accuracy as its six-week trend, and hold performance to baseline in the register as the lagging pair. Gap to target surfaces slip as it forms; the other two score delivery discipline after the fact. None of it can be published until one plan of record exists [18].

Quality: one number, two businesses

As a single line, product quality lets unrelated problems hide behind one number. On one of this client's high-volume product families, returns that test within specification account for more than half of all returns; on another, the dominant mode is a genuine hardware failure [17]. Those are different businesses. The first is a usability, expectation-setting and documentation problem; the second is a reliability defect. One headline number reports them identically, and the corrective actions have nothing in common. Field quality should therefore be reported as the returns rate by SKU with test-in-spec separated out. Ecosystem quality (pairing success, connection drops, data synchronization loss, update success, application crashes) does not exist in the current measurement set at all, and it is the category that will define the brand once the products connect. A technically perfect product that will not pair reliably is a defective product to the person holding it. In the two-clock system both quality lines are reviewed at the gates and the monthly quality review; the weekly page carries the leading indicators of the date. The deadline holds regardless: ecosystem telemetry cannot be retrofitted onto shipped hardware, so it has to be specified as a first-release product requirement.

Break-even: the metric that makes a team a business

Time to break-even is the strongest metric on the proposed list. House and Price described the practice at Hewlett-Packard in 1991: putting break-even time on the wall makes a cross-functional product team behave like a business, because a team that owns the product through to the point it makes money makes different design decisions, earlier, than a team that hands off at manufacturing transfer [4]. Keep it, keep the ownership with the product team rather than with finance, and give software a separate definition, customer-acquisition-cost payback, so one forced definition does not kill the software investment or corrupt the hardware metric. The metric lives where House and Price put it, with the team, set alongside cost of delay at the funding gate; portfolio yield asks the same question quarterly at portfolio level.

Note on pairing. Every speed metric needs a counterweight or it produces the behavior it was meant to prevent [8]. Gap to target alone produces padded plans; it is paired with milestones achieved to plan on the same weekly page, with forecast accuracy read as its trend. Speed alone produces cut corners; the quality lines stand behind the weekly page at the gates. Decision velocity alone produces hasty calls; it is paired with decision round-trips reviewed at the gate. No metric in this framework is published without its pair.

Principle

The measurement problem

**“What you
measure is
what you get.”**

Robert Kaplan and David Norton
Harvard Business Review, January–February 1992

04

The set **Ten numbers, on two clocks**

A set of metrics does not become a control system until each number is owned by a body that can act on it, at a cadence matched to how fast the number can move. That last clause is the design insight. Execution numbers and business numbers move at different speeds: put them on one page at one cadence and they produce a report; split them onto two clocks and they produce two different decisions, namely what we change this week, and whether the strategy is working. The client's operating model already defined the two audiences: a portfolio committee governing by exception, and a program tier run by product development leadership with the CEO, with roughly ten product teams beneath them, each led by a triad of product, technical and project owners, holding about eighty percent of day-to-day decisions.

Metric	Owner	Cadence	Target	Practice it measures
The weekly clock — run the program · CEO and VP Product Development				
Gap to target Bottom-up critical path against committed date, in weeks, per team	Program Manager	Weekly	Zero, or escalated	Weekly schedule refresh; targets and trends
Critical-path role fill Core roles named and released on every committed team	VP PD / functional leads	Weekly	100% on committed teams	Dedicated core team
Decision velocity Median days from decision identified to decision made	Leadership Triads	Weekly	Team under 2 days; program under 7	The bounding box
Provisioning and interrupts Request-to-release latency; host interrupts, trailing 12 months	Functional leaders	Weekly	10 days or less; zero interrupts	Host behavior
Team health Eight operating criteria per team, published worst first, never averaged	PMO Director	Weekly	Above 90%	Aggregate, then drill down
Milestones achieved to plan Phase gates and releases passed year to date against plan	Project Owners	Weekly	100%	Rhythm of accountability
The quarterly clock — run the portfolio · Portfolio committee				
Portfolio yield Annual incremental growth above run-rate over active portfolio NPV	VP PD with Finance	Quarterly	5%	Right-time financial return
Innovation vitality Revenue from products launched inside a 30-month window	Finance / Category Directors	Quarterly	30%	Portfolio refresh
Customer evidence coverage Active programs with the full evidence set before development commit	Upstream Marketing	Quarterly	100%	Manage the fuzzy front end
Strategic capability maturity Assessed maturity against the approved capability roadmap	VP PD with functional leaders	Quarterly	Roadmap milestones	Build the muscles the strategy assumes

Figure 4. The ten. Every metric is a measurement placed on top of an FTTM practice the organization is implementing; the practice is the lever that moves the number. The cadence column is the refresh rate of the number; the clock header is the forum that reviews it.

Each clock gets numbers it can act on

Mapping metrics onto the governance structure resolves the most common failure in corporate scorecards, which is sending every number to every audience. The weekly page belongs to the CEO and product development leadership: six numbers, every one of which can change inside a week. The quarterly page belongs to the portfolio committee: four numbers that move at business speed. The teams keep their own daily boards beneath both. A committee should not receive weekly critical-path detail it has no mechanism to act on, and a product team should not be scored on portfolio yield it cannot influence.

One clarification prevents the whole framework from being misread. Every metric in Figure 4 has a single named owner, and that can look like the functional scorecard returning through the side door. Owner here means steward, not sole proprietor: the named role defines the number, publishes it, and brings its cause and corrective action to the room. Accountability for what the number says belongs to the leadership team as one body. The distinction is the framework: a scorecard where each owner also owns the outcome is five functional scorecards stapled together, and the mutual accountability that Katzenbach and Smith put at the center of real teams never forms [13].

Why ten

Ten is a ceiling chosen in advance, not a count of what survived. The working-memory literature, from Miller's "magical number seven" to the stricter modern estimates, is a caution against dashboards of thirty rather than a formula for ten [22, 23]; the operational argument is the one that matters. A management team can hold roughly a dozen numbers across a quarter and still know what each one means, who owns it, and what would cause it to move. Past that, the review becomes a reading exercise: attention spreads evenly across measures of wildly unequal importance, and the meeting produces commentary instead of decisions. Six and four is an outcome of the two clocks, not numerology. The selection rule is what carries the weight: a metric earned a place only if the management team would change a decision because of it. Even the most conventional entry passes that test: milestones achieved to plan exists to confirm, week after week, that the five leading indicators above it are telling the truth. Everything else went into the supporting register described in section 06.

05

The clocks

The weekly six and the quarterly four

The two clocks answer two questions. The weekly six: will we hit the date, and what do we change this week? The quarterly four: is the strategy working? This section walks both, with the FTTM practice each number measures and the specific failure it is designed to catch. The association with practice is the point. A metric with no practice behind it is a complaint; a metric attached to a practice tells you which lever to pull.

The weekly six: run the program

Metric	Practice it measures	Evidence from the engagement
Gap to target	The weekly schedule refresh, read as targets and trends: the target stays fixed, the forecast tells the truth, and the trend between them is the management information.	First reading: +9 weeks. Two planning views disagreed on later scope, and no rolled-up critical path existed.
Critical-path role fill	The dedicated core team. Filled means named by the functional manager and released; a name with no released time counts as zero.	17 of 56 core roles filled, a 30 percent reading. Industrial design, quality and test unfilled at program level.
Decision velocity	The bounding box: with decision rights defined, about 80 percent of decisions belong inside the team.	First reading: 12 days median. Three interviewees described decisions being punted upward.
Provisioning and interrupts	Host behavior. Five of the six measure the teams; this one measures leadership.	34-day median from request to release. Two late reversals cost months and appeared in no number leadership reviewed.
Team health	Aggregate, then drill down: every team against eight criteria, worst first, never averaged.	A portfolio average would have read 46 percent while two teams sat at 13.
Milestones achieved to plan	The rhythm of accountability, and the lagging check on the five above it.	First reading: 71 percent of gates and releases passed to plan.

Figure 5. The weekly six: the practice each number measures, and the engagement finding behind it [17].

Five of the six are leading indicators, which is the point of the clock. Gap to target moves before a milestone is missed. Decision velocity moves before a program stalls; the return on fast decisions is one of the most reliable findings in the lateralworks field research [24]. Role fill moves before a team fails to form, provisioning latency moves before the host becomes the constraint, and team health moves before the failure reaches the schedule. Only milestones achieved to plan is retrospective, and it earns its place as the honesty check: it is the number that says whether the five leading indicators are being gamed. Every red on the page arrives with its root cause and corrective action already written, once, not weekly.

The quarterly four: run the portfolio

Portfolio yield asks whether the portfolio as a whole is producing new revenue above run-rate for the NPV it carries: the right-time discipline of treating product development as a financial instrument [27]. Innovation vitality reports whether the portfolio is refreshing, value-weighted and hard to game; it carries the 3M pedigree described in section 03 [14], counted here inside a 30-month window. Customer evidence coverage counts the share of active programs that entered development with the full evidence set behind them, assessed before the commit: the fuzzy front end made measurable [19, 20]. At this client the first reading of 38 percent was the number behind the lead program's rework, and the metric exists so that does not happen twice [17]. Strategic capability maturity tracks whether the organization is building the muscles the strategy assumes it has, assessed against an approved roadmap rather than against ambition.

These four move too slowly to manage against and too consequentially to ignore, which is exactly why they get their own clock. Reviewed weekly they would train the room to skip them; reviewed quarterly, by the

body that owns starts, pauses and kills, they are the standing answer to whether the strategy is working.

The connective number

Cost of delay is the exchange rate between the two clocks. Each team carries its own figure, in dollars per calendar day, set at the funding gate, refreshed when the business case changes, and published to the team rather than held at portfolio level [21, 25, 26]. It converts a schedule gap into a sum of money the company is currently choosing to spend, prices every decision that ages past its target, and turns every pull-in action into an investment case with a return. Without it, acceleration is a favor someone does for the program office. With it, buying back a day with a contractor or a second test rig is an ordinary financial decision made in the room.

06

The cut

What is deliberately excluded, and why

A framework that lists only what it includes invites the same argument every quarter, because each function reasonably assumes its own measure was overlooked rather than considered. So the framework specifies the excluded measures and deliberately holds them in a register, each with a stated reason and a route to promotion. The register also protects the ten: a headline set with no visible discipline behind it is indistinguishable from an arbitrary one, and arbitrary sets grow.

Four reasons a measure stays off the wall

Reason	#	Measures held in the register	Where they surface
Working diagnostic	5	Schedule movement; float consumed; aged open decisions; requirements volatility; team dedication rate.	The teams' daily boards and weekly packs. Rolls into gap to target, decision velocity and role fill.
Sub-component or lagging pair	5	Forecast accuracy (the six-week gap trend); first-pass gate yield; performance to baseline; warranty and returns cost; launch confidence.	Read from the same instruments, or reported at the gate as supporting detail.
Measured at its own rhythm	7	Market share; connected attach rate; recurring revenue; field quality; ecosystem quality; cycle time; time to break-even.	The gates, the monthly quality review, and the annual strategy cycle. Attach rate and recurring revenue go live at the first connected release, which is what makes the telemetry decision urgent.
Gate measure, not a trend	5	Cost of delay; kill rate; development quality; owned-IP share; customer outcome.	The funding gate, phase-gate reviews and the program charter. Customer outcome is the promotion candidate.

Figure 6. The supporting register, grouped by reason. Nothing is discarded; each measure has a defined home and a route back.

Three of the four reasons are structural rather than judgments about importance. A working diagnostic does not become a company metric by being useful. A gate measure loses meaning when forced onto a weekly rhythm. And a number that cannot yet be measured, like attach rate before telemetry exists, earns its way onto a wall by becoming measurable, not by being wished there. Only the sub-component category involves a real trade-off, and in each case the parent metric is the one that would trigger action first. Warranty cost is the standing example: it is the number the P&L; feels, and it moves only when field quality moves, so it is reported as the standing financial footnote to the quality review rather than as a competing line.

The rule for promotion

Rule for promotion. A register metric moves onto the headline set only by displacing one that is already there, and only at the annual review. The portfolio committee approves the swap and records the reason. Without a fixed size, every good argument for adding a metric wins, and the set returns to thirty-two within two years.

The register is maintained alongside the plan of record, reviewed at each phase gate for the programs in scope, and reviewed in full at the annual strategy cycle. Every entry carries the same fields as a headline metric (definition, owner, source system, cadence, target), so promotion is an administrative act rather than a fresh design exercise. A measure that is not maintained in the register cannot be promoted, which is the mechanism that keeps the ten from being expanded by whichever function argues hardest in the room.

07

The wall

The one-page dashboard

A measurement framework lives or dies on how it is consumed, and the consumption format here is deliberately physical: a single page, designed to be printed large and hung in the room where the portfolio committee meets. Figure 7 reproduces it, anonymized, exactly as built for the client. One page carries all ten with owner, cadence, target and the FTTM practice each one measures, plus panels for the connective number and what is deliberately excluded, so the discipline of the cut hangs on the wall next to the numbers it protects.

The point of one page is not graphic economy. It is that a scorecard compact enough to be seen whole can be reviewed by the leadership team as one body. The alternative, each executive reviewing their own numbers in their own meeting, reproduces exactly the vertical measurement the framework exists to replace.

The dashboard is consumed on three rhythms, each mapped to an existing meeting rather than a new one. Weekly, the CEO and product development leadership run the six operating numbers from the drill-down views. Monthly, the portfolio committee reviews by exception. Quarterly, it takes the business four, and the full leadership team stands in front of all ten on the one page. Beneath all three, every team runs its own daily board. Nothing new is scheduled; the existing reviews change what they look at.

› › The performance dashboard Ten numbers on two clocks. Every one has a named owner, a named source, and an agreed definition before it is published. Six numbers reviewed weekly to change the date. Four reviewed quarterly to judge the strategy. Every red arrives with its cause and corrective action written. Client identity withheld August 2026 · Rev 2.0			
Metric	Owner	Cadence	Practice it measures
The weekly clock — run the program · CEO and VP Product Development · weekly			
Gap to target Bottom-up critical path against committed date, in weeks, per team	Program Manager	Weekly	Weekly schedule refresh; targets and trends
Critical-path role fill Core roles named and released on every committed team	VP PD / functional leads	Weekly	Dedicated core team
Decision velocity Median days from decision identified to decision made	Leadership Triads	Weekly	The bounding box
Provisioning and interrupts Request-to-release latency; host interrupts, trailing 12 months	Functional leaders	Weekly	Host behavior
Team health Eight operating criteria per team, published worst first, never averaged	PMO Director	Weekly	Aggregate, then drill down
Milestones achieved to plan Phase gates and releases passed year to date against plan	Project Owners	Weekly	Rhythm of accountability
The quarterly clock — run the portfolio · Portfolio committee · quarterly			
Portfolio yield Annual incremental growth above run-rate over active portfolio NPV	VP PD with Finance	Quarterly	Right-time financial return
Innovation vitality Revenue from products launched inside a 30-month window	Finance / Category Directors	Quarterly	Portfolio refresh
Customer evidence coverage Active programs with the full evidence set before development commit	Upstream Marketing	Quarterly	Manage the fuzzy front end
Strategic capability maturity Assessed maturity against the approved capability roadmap	VP PD with functional leaders	Quarterly	Build the muscles the strategy assumes
The connective number Cost of delay, in dollars per day per team, converts every gap and every late decision to money. Set at the funding gate, published to the team, refreshed when the business case changes. Forecast accuracy is the six-week trend on gap to target; one instrument, two views.		Deliberately not on this wall The team operating rows. Critical path, pull-in actions, decisions, commitments and impediments are the daily board inside every team. They feed gap to target, decision velocity and provisioning above. The supporting register. Market share, attach rate, recurring revenue, field and ecosystem quality, cycle time, break-even, gate yield, warranty cost and the rest are reviewed at the gate, the monthly quality review, or the annual strategy cycle.	

Full specification: Ten numbers, Rev 2.0

© 2026 lateralworks

Figure 7. The one-page dashboard, reproduced from the engagement with client identifiers removed. Ten metrics, two clocks, one page: rotate the page to read, or stand in front of the wall-chart version. [17]

Run the meeting from the roll-up

The dashboard's first metric, gap to target, is deliberately a roll-up. Behind the single number per release sits a portfolio view: every program, its target milestone, the bottom-up finish date, the gap in days, the six-week trend, the cause of any change, and the actions in place. That structure lets the leadership team use the dashboard dynamically. The operative question is always some form of “which milestones are late and slipping, and what are we doing about it?” — and because cause of change and actions are filled in before the meeting, the answer is on the page, not in somebody's follow-up [18, 28].

Program	Target milestone	Gap (days)	6-week trend	Cause of change	Actions in place
Program A	Release to production	(37)	(22) and opening	Problems with the latest prototype build; one process step under benchmark.	Removal of the failing step from the flow to be confirmed; qualification lots tracked to the revised plan; owners and dates named per action.

Figure 8. One anonymized row of the drill-down behind gap to target. Parentheses denote days late; the trend column shows the average movement over the past six weeks and whether the gap is opening or closing. Every red gap arrives with its cause and its actions already written [17].

This is the lateralworks aggregate-then-drill-down pattern: roll every program's schedule trend up to one number, filter for the largest gaps, and drill into root cause only where the number says to [29]. Cadence matters as much as content: a review that meets every Monday, starts on time and never postpones builds the rhythm of accountability that makes trend data mean something [28, 30]; a review that floats produces trend charts with holes in them.

The weekly discipline behind the wall

The weekly page runs on a strict red discipline. Its two liveliest rows, gap to target with milestone trend and critical-path role fill, carry the discussion; the others show their last reading with its date. The discipline that makes it work is the treatment of red: every red arrives with a complete root-cause analysis and corrective action in the next Monday pack, once, not weekly. While a row stays red it shows the recovery metric and the recovery date, and that is the discussion. A missed recovery date reopens the analysis, and the reopening is the agenda item [17].

Note what the red discipline is not: it is not tighter control. The lateralworks research on control systems found that high-control organizations were not the fast ones; the fast ones paired autonomous teams with honest, frequent, low-ceremony measurement [31]. A red on this wall triggers analysis and help, stated as a system failure, not a person failure. The moment a red triggers punishment, the organization stops producing honest reds, and the dashboard turns green and useless in the same quarter.

One practice completes the system: the leadership team scrubs the ten together, as one standing agenda, rather than each executive reviewing their own metrics with the CEO one-to-one. The set is cross-functional; reviewing it functionally would convert it back into five separate scorecards. Katzenbach and Smith's definition of a real team — mutual accountability to common performance goals — is, in the end, a statement about what gets reviewed in whose presence [13].

08

Path forward

Making the numbers real

Most measurement systems fail on definition rather than collection. The numbers get published, two functions calculate them differently, the review turns into an argument about the denominator, and within three quarters the dashboard is decoration. This closing section lists the definitional decisions that make the ten real, the build sequence, and the decisions a leadership team has to take in the room.

Six definitions to settle before measurement starts

#	Decision required	Recommendation
1	Which baseline counts, and the re-baseline rule	The funding-gate baseline. Re-baselining only at a gate, only with committee approval, and the original commitment stays visible.
2	What counts as a decision, and when its clock starts	Logged when first identified, closed when recorded; tracked separately for product, technical and program decisions.
3	What “filled” means for role fill	Named by the functional manager and released to the team. Named without released time counts as zero.
4	What counts as an interrupt	Late reversals, re-prioritizations, and resources pulled off a critical path. The provisioning clock starts at the team’s request, not the manager’s acknowledgment.
5	What counts as a launch for the vitality index	A distinct product family launch, not a SKU. Derivatives and variants do not reset the clock.
6	What evidence counts for coverage	All six elements, from target definition through validated jobs-to-be-done outcomes to an approved requirements document, assessed before the development commit.

Figure 9. Definitional decisions required before publication. Each blocks a named headline metric.

The first item is the constraint on the whole weekly clock. Gap to target is unmeasurable until the competing planning views are reconciled into one plan of record with a rolled-up critical path. Publishing any schedule number before that exists produces a figure each function can dispute on its own terms, which is worse than publishing nothing. The sixth item decides whether the coverage number means anything: count partial evidence as covered, and the metric will read comfortably while programs keep repeating the rework it exists to prevent.

Sequence the build around what exists

The build sequence respects data reality rather than framework elegance, because launching ten metrics at once across eight teams produces estimated numbers, and estimated numbers on a CEO scorecard teach the organization to produce comfortable ones. Weeks one through four build the instrument: reconcile the timeline sources into one plan of record, on a thirty-day clock, with the program office owning the work and the portfolio committee arbitrating scope disputes; stand up the weekly refresh on every team; open the decision log; and add the impediment row to every daily board. The decision log is deliberately minimal, four fields kept from existing gate and committee minutes (decision, date identified, deciding body, date resolved); a spreadsheet, not a system. Weeks five through eight produce the first measured pack: gap to target, role fill and milestones go live as measured values, decision velocity and provisioning latency produce their first real readings, team health publishes worst first, and the weekly pack to the CEO begins. The following quarter completes the system: the quarterly four reviewed with finance-confirmed figures, per-team cost of delay agreed and published, targets reset off measured baselines, and the system extended beyond the lead program to the wider portfolio. Until a number is measured, the pack marks it as an estimate, without exception; publishing the first honest readings early is what establishes the review rhythm before the harder numbers arrive [28].

Six decisions for the room

Confirm the ten numbers and the two clocks. Six weekly, four quarterly. Everything else moves to supporting detail. Name an owner for every number; until a role is filled, the clock's governing body holds it and the fill date appears on the dashboard itself.

Approve provisioning latency and the interrupt count. This metric measures leadership, not the teams, and it needs explicit sponsorship from the top of the room to survive its first red.

Sign the per-team cost-of-delay figures with Finance. Allocated from the R&D; strategy numbers and published to the core teams, not held at portfolio level.

Set the forecast tolerance. The gap trend needs an agreed variance band before the first pack; a band of plus or minus two weeks once a baseline exists is the starting proposal.

Take the staffing decision on the worst unit this quarter. Thirty-two weeks behind with no team named is not a schedule problem. Build it or pause it.

Accept estimates on the first two packs. The numbers become real in week five. Punishing the first honest reading is how the organization learns to produce comfortable ones after that.

The compensation question

Kerr's folly, hoping for cross-functional cooperation while paying for functional results, does not end until pay moves, and the framework takes a deliberate position on when. Not in year one: linking compensation to unsettled definitions starts the gaming before the measuring, for the same reason a punished red stops being an honest red [6]. At the first annual review, once the definitions have survived four quarters, tie a common slice of every executive's variable compensation to the same quarterly four: the same numbers, at the same weight, for every member of the team. Individual functional incentives can continue underneath; the shared slice is what makes the ten everyone's problem.

The close

Functional measurement is not the enemy; it is simply insufficient, and it is what organizations default to because functions are how the org chart is drawn and how careers are paid. The company scorecard is the one instrument a leadership team fully controls that reaches across that default. Point it at the horizontal flow, at the ten numbers that say whether the date will hold and whether the strategy is working, and review it as one team, and the measurement system starts pulling the organization toward the thing it says will grow it. The scorecard is a strategy document. Write it like one.

Sources

References

- [1] Manly, J., Ringel, M., MacDougall, A., et al. "Innovation Systems Need a Reboot." Boston Consulting Group, Most Innovative Companies 2024, June 2024. <https://www.bcg.com/publications/2024/innovation-systems-need-a-reboot>
- [2] Knudsen, M. P., von Zedtwitz, M., Griffin, A., and Barczak, G. "Best practices in new product development and innovation: Results from PDMA's 2021 global survey." *Journal of Product Innovation Management*, Vol. 40, No. 3, 2023, pp. 257-275.
- [3] Oregon Encyclopedia. "Tektronix, Inc." Oregon Historical Society. https://www.oregonencyclopedia.org/articles/tektronix_inc/
- [4] House, C. H., and Price, R. L. "The Return Map: Tracking Product Teams." *Harvard Business Review*, Vol. 69, No. 1, January-February 1991, pp. 92-100.
- [5] Ridgway, V. F. "Dysfunctional Consequences of Performance Measurements." *Administrative Science Quarterly*, Vol. 1, No. 2, 1956, pp. 240-247.
- [6] Kerr, S. "On the Folly of Rewarding A, While Hoping for B." *Academy of Management Journal*, Vol. 18, No. 4, 1975, pp. 769-783. Republished in *Academy of Management Executive*, Vol. 9, No. 1, 1995.
- [7] Kaplan, R. S., and Norton, D. P. "The Balanced Scorecard - Measures That Drive Performance." *Harvard Business Review*, Vol. 70, No. 1, January-February 1992, pp. 71-79.
- [8] Hauser, J. R., and Katz, G. M. "Metrics: You Are What You Measure!" *European Management Journal*, Vol. 16, No. 5, 1998, pp. 517-528.
- [9] lateralworks. "Integrated Core Team." [lateralworks.com/ideas](https://lateralworks.com/ideas/integrated-core-team), January 2019.
- [10] Cooper, R. G., Edgett, S. J., and Kleinschmidt, E. J. "Benchmarking Best NPD Practices I-III." *Research-Technology Management*, Vol. 47, Nos. 1, 3 and 6, 2004.
- [11] Cooper, R. G. "Pathways to Profitable Innovation." Product Development Institute working paper summarizing the APQC benchmarking study, 2005.
- [12] Ancona, D. G., and Caldwell, D. F. "Demography and Design: Predictors of New Product Team Performance." *Organization Science*, Vol. 3, No. 3, 1992, pp. 321-341.
- [13] Katzenbach, J. R., and Smith, D. K. "The Discipline of Teams." *Harvard Business Review*, Vol. 71, No. 2, March-April 1993, pp. 111-120.
- [14] Hindo, B. "At 3M, A Struggle Between Efficiency and Creativity." *BusinessWeek*, June 11, 2007.
- [15] The Motley Fool. "The 3M You Don't Know." June 26, 2013. <https://www.fool.com/investing/general/2013/06/26/the-3m-you-dont-know.aspx>
- [16] Gordon, M., Kowski, M., and Smits, S. "Taking the measure of product development." McKinsey & Company, Operations Practice, October 2018.
- [17] lateralworks. Internal engagement records, performance measurement engagement, 2026. Client identity withheld.
- [18] lateralworks. "Targets and Trends." [lateralworks.com/ideas](https://lateralworks.com/ideas/targets-and-trends), January 2019.
- [19] lateralworks. *FTTM New Product Development Best Practices*, Revision 2018.002. Sections 1.6, 2.1, 2.4, 2.5 and 2.9.
- [20] lateralworks. "Manage the Fuzzy Front End." [lateralworks.com/ideas](https://lateralworks.com/ideas/manage-the-fuzzy-front-end), January 2019.
- [21] Smith, P. G., and Reinertsen, D. G. *Developing Products in Half the Time: New Rules, New Tools*, 2nd ed. John Wiley & Sons, 1998.

- [22] Miller, G. A. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." *Psychological Review*, Vol. 63, No. 2, 1956, pp. 81-97.
- [23] Cowan, N. "The magical number 4 in short-term memory: A reconsideration of mental storage capacity." *Behavioral and Brain Sciences*, Vol. 24, No. 1, 2001, pp. 87-114.
- [24] lateralworks. "What is the R.O.I. of fast decision-making?" lateralworks.com/ideas.
<https://lateralworks.com/ideas/what-is-the-roi-of-fast-decision-making>
- [25] lateralworks. "Cost-of-Delay." lateralworks.com/ideas, January 2019. <https://lateralworks.com/ideas/cost-of-delay-2>
- [26] Reinertsen, D. G. *The Principles of Product Development Flow: Second Generation Lean Product Development*. Celeritas Publishing, 2009.
- [27] lateralworks. "Right-time ensures financial return from new products." lateralworks.com/ideas, June 2019.
<https://lateralworks.com/ideas/right-time-ensures-financial-return-from-new-products>
- [28] lateralworks. "The Rhythm of Accountability." lateralworks.com/ideas, January 2019.
<https://lateralworks.com/ideas/the-rhythm-of-accountability>
- [29] lateralworks. "Aggregate, then Drill Down." lateralworks.com/ideas, January 2019.
<https://lateralworks.com/ideas/aggregate-then-drill-down>
- [30] lateralworks. "The Weekly Schedule Refresh." lateralworks.com/ideas.
<https://lateralworks.com/ideas/the-weekly-schedule-refresh>
- [31] lateralworks. "Do tighter control systems create faster projects?" lateralworks.com/ideas, June 2019.
<https://lateralworks.com/ideas/do-tighter-control-systems-create-faster-projects>