Whitepaper

Twelve 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 twelve-metric framework, 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

Twelve numbers

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 twelve numbers 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 twelve metrics in three groups of four: enterprise outcome, reviewed quarterly by the CEO and the portfolio committee; portfolio and program health, reviewed monthly; and quality and economics, owned by product owners, quality and finance. Four of the five survive, each sharpened; the fifth is replaced. The engagement is hardware, but the method, and most of the twelve, 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 twelve, with twenty further measures held in a register, each with a stated reason for exclusion and a rule for promotion. And the twelve 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, four of the twelve could be published within a quarter from data already sitting in the phase-gate record, the returns database and the finance system. Five more needed sixty days and no new systems or 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.

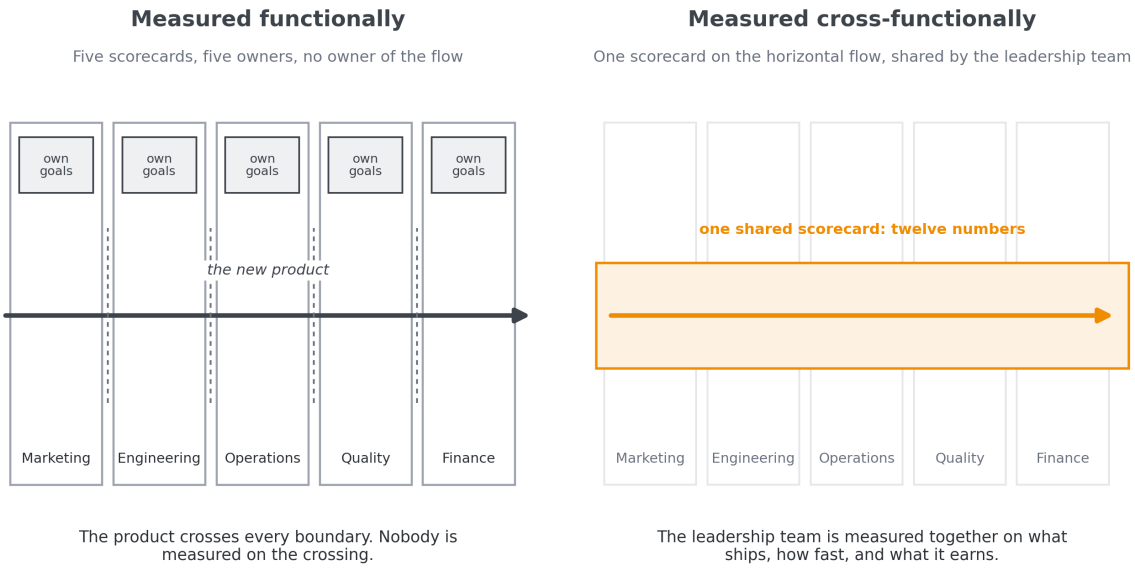

Figure 1. The same five functions, measured two ways. A functional scorecard puts a number on every vertical and none on the flow that generates the growth. A cross-functional scorecard measures the leadership team, together, on the flow.

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 twelve that replace them, the discipline of what stays off the wall, and the one-page dashboard that puts the twelve 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	Keep	Start the clock at first idea discussion, not at team formation. Split into gestation, development, and ramp.
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	Publish gap to target as the headline. Performance to baseline becomes its lagging pair in the register.
4	Product quality	Keep, split	Separate field quality from ecosystem quality. One number hides two unrelated problems with opposite fixes.
5	Time to break-even	Keep	Define a separate subscription equivalent for software. Cost of delay becomes the before-the-fact companion, set at the gate.

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 fix is to split the measure into gestation, development and ramp, and to normalize by program class so the year-over-year comparison means something.

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 headline set, and hold performance to baseline behind it as the lagging pair. Gap to target surfaces slip as it forms; performance to baseline scores delivery discipline after the fact. A management team needs the first every week and the second every quarter, and it can have neither 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.

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.

> **Note on pairing.** Every speed metric needs a counterweight or it produces the behavior it was meant to prevent [8]. Cycle time alone produces cut corners; it is paired with field quality and business case forecast accuracy. Gap to target alone produces padded plans; it is paired with performance to baseline in the register. Adoption alone produces channel stuffing; it is paired with retention inside the recurring revenue line. 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

The twelve, mapped to the operating model

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. The client's operating model already defined three levels of decision-making: a portfolio management committee governing by exception, a program tier owning end-to-end delivery, and roughly ten product teams, each led by a triad of product, technical and project owners, holding about eighty percent of day-to-day decisions. The measurement framework follows that structure rather than inventing a parallel one.

Metric	Owner	Cadence	Target	Live from
Tier 1 — Enterprise outcome · CEO and PMC · quarterly				
Vitality index Revenue from products launched in the last three years	CFO / Category Directors	Quarterly	20%	Now
Market share by category Unit and revenue share per category, by region	Category Directors	Quarterly	Defend share in the two lead categories	Now
Connected attach rate Products activated and paired within 30 days of registration	Digital lead	Monthly	Set at Release 1	R1 launch
Recurring revenue ARR, with paid conversion and net revenue retention	Digital lead	Monthly	Against the five-year ramp	First revenue
Tier 2 — Portfolio and program health · PMC · monthly				
Gap to target, per release Bottom-up finish date against committed release date	Program Manager	Weekly	Zero, or escalated	60 days
Decision latency Days from decision requested to decision recorded, by deciding body	EPMO	Monthly	PMC under 14 days; triad under 3	60 days
Critical-path role fill rate Core roles named and allocated on every committed team	Program Manager	Monthly	100% on committed programs	60 days
Business case forecast accuracy Actual 12-month revenue against the gate-approved forecast	EPMO	Quarterly	Within 20%	60 days
Quality and economics · Product Owners, Quality and Finance				
Field quality Rolling twelve-month returns rate by SKU, test-in-spec separated	Quality lead	Monthly	At or below current line rates	Now
Ecosystem quality Pairing, connection, data sync, update and app crash rates	Software Technical Owner	Weekly	Set at Release 1	R1 launch
Cycle time Gestation, development and ramp, measured as three segments	EPMO	At closure	Reduction against class baseline	60 days
Time to break-even Months to positive contribution; CAC payback for software	Product Owner / Finance	Post-launch	Against the gate business case	Now

Figure 4. The twelve. “Live from” marks when each becomes measurable: now, sixty days, at first connected release, or at first subscription revenue. The cadence column is the refresh rate of the number; the tier header is the cadence of the forum that reviews it. EPMO: enterprise program management office.

Each audience gets four numbers it can use

Mapping metrics onto the governance structure resolves the most common failure in corporate scorecards, which is sending every number to every audience. A product team should not be scored on recurring revenue it cannot influence. The portfolio committee should not receive weekly critical-path detail it has no mechanism to act on. Four numbers per audience, each one actionable by that audience, is the design rule.

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 twelve

Twelve 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 twelve [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. The grouping into three fours is a design convenience. The selection rule is not: 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: market share by category exists to force a quarterly portfolio-balance decision about defending the lead categories the new platform is meant to protect. Everything else went into the supporting register described in section 06.

05

The groups

Three groups of four

The three groups answer three different questions. Enterprise outcome: are we becoming the company the strategy describes? Portfolio and program health: will we deliver what we have committed, and can we see trouble early? Quality and economics: is what we ship good, and does it make money? This section walks each group of four, with the specific failure it is designed to catch.

Enterprise outcome: is the strategy happening?

The four Tier 1 metrics map one-to-one onto the strategic priorities, so the quarterly review reads as a direct answer to the commitments the company has made to itself. The vitality index replaces products per year and reports whether the portfolio is refreshing. Market share by category defends positions the new platform is meant to protect: when a company replaces nearly its whole product line over three years, the downside case is losing share it already holds. Connected attach rate, products activated and paired within thirty days, is the single best leading indicator of a platform thesis, because hardware sold is not ecosystem built, and this number diverges from unit sales long before revenue shows it. Recurring revenue, with paid conversion and net revenue retention inside it, attaches a measurement to what is otherwise a set of assumptions with a ramp attached.

One strategic priority, technology leadership, deliberately has no headline metric. Its candidate measures, owned-IP share of shipped product value and platform reuse rate, move on an annual cycle, and a number reviewed four times a year that only moves once a year trains the room to skip it. Both sit in the register on an annual cadence, with the open question of promotion recorded in section 08.

One honest gap should be named rather than papered over. The five were faulted because none would move if the strategy died. Yet until the first connected release ships, the two metrics that watch the digital strategy are dark, and the vitality index trails by construction. In year one, the strategy is measured by proxies: whether the telemetry requirement lands in the release contract, and gap to target on the release that carries the strategy. That is exactly why the ecosystem-telemetry decision in section 08 cannot be allowed to slip; it is the moment the strategy becomes measurable at all.

Portfolio and program health: see trouble early

Metric	What it detects	Evidence from the engagement
Gap to target, per release	Bottom-up reality against the committed date. Surfaces slip as it forms rather than at the next gate.	The first release had already moved a quarter; two planning views disagreed on later scope; no rolled-up critical path existed.
Decision latency	Days from decision requested to decision recorded. Makes decision speed a leadership metric rather than a complaint.	Decision cadence was the single named environmental weakness. Three independent interviewees described decisions being punted upward.
Critical-path role fill rate	Committed programs running without named owners.	One program carried a launch commitment two years out with no team named across all seven core roles; another was missing its product and technical leads.
Business case forecast accuracy	Actual 12-month revenue against the gate-approved forecast. Whether capital is allocated on analysis or on theater.	The quality policy already required the comparison at post-project review. Nobody trended it across the portfolio.

Figure 5. Tier 2 metrics, with the engagement-specific failure each one is designed to catch [17].

Three of the four are leading indicators, which is the point of the group. Gap to target moves before a milestone is missed. Decision latency 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 rate moves before a team fails to form. Only forecast accuracy is retrospective, and it earns its place as the honesty metric: it is the one number that says whether the analysis underneath every other investment decision can be trusted. It will be

uncomfortable in its first year, and that should be said in the room before it is published. Forecasts written under commercial pressure to clear a gate are not forecasts. The metric will show that, and the correct response is to change how business cases are built, not to change the metric.

Quality and economics: good, and paid for

Field quality and ecosystem quality carry the split argued in section 03. Cycle time, in its three segments, reports whether the engine is speeding up. Time to break-even reports whether the products the engine ships earn their development back, which is where the financial return from a fast engine ultimately lands [27]. Two companions sit just behind this group in the register. Cost of delay, contribution lost per day of launch delay, is not a trend; it is a standing figure calculated once at the funding gate, refreshed when the business case changes, and published to the team [21, 25, 26]. Its purpose is to change what a resourcing conversation sounds like when a program is three weeks behind and one hire would recover two of them. Customer outcome, the improvement the product actually delivers in the field, is the number the whole value proposition rests on, and it becomes measurable about two quarters after the first connected release. It is the strongest candidate for promotion into the headline set.

Ecosystem quality carries the only irreversible deadline in the framework. Pairing success, connection drops, data loss, update success and crash rates cannot be retrofitted onto shipped hardware. The instrumentation has to be specified as a first-release product requirement, with a defined telemetry contract at the boundary where device data enters the cloud platform. This is a product decision with a measurement consequence. Miss it, and the company launches a connected ecosystem with no ability to see whether it is working — the first reliable signal becomes the returns rate six months after launch, by which point the next release’s architecture is already committed.

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 twenty further measures and deliberately holds them back, each with a stated reason and a route to promotion. The register also protects the twelve: 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 product teams' weekly pack. Rolls into gap to target, decision latency and role fill rate.
Sub-component or lagging pair	6	Connected products per account; usage per active customer; warranty and returns cost; portfolio load against capacity; first-pass gate yield; performance to baseline.	Drill-down on exception when the parent metric moves.
Annual cycle	3	Owned-IP share of product value; platform reuse rate; platform versus product spend.	Annual strategy review and portfolio planning.
Gate measure, not a trend	6	Development quality; kill rate; cost of delay; cross-team dependency health; days lost to shared functions; customer outcome.	Phase-gate review and program charter. Customer outcome is the promotion candidate.

Figure 6. The twenty excluded measures, grouped by reason. Nothing is discarded; each has a defined home.

Three of the four reasons are structural rather than judgments about importance. A working diagnostic does not become a company metric by being useful. An annual measure does not improve by being read quarterly. A gate measure loses meaning when forced onto a monthly rhythm. 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 permanent financial footnote beneath field quality 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 twelve 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 twelve metrics with owner, cadence, target and availability, plus a panel naming 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 program tier runs gap to target and ecosystem quality from the drill-down views, alongside a six-row scorecard that carries the live conversation with the CEO. Monthly, the portfolio committee reviews the middle and lower tiers by exception. Quarterly, the full leadership team stands in front of all twelve on the one page. Nothing new is scheduled; the existing reviews change what they look at.

The performance dashboard

Twelve metrics. Every one has a named owner, a named source, and an agreed definition before it is published.
Each metric carries a counterweight: speed is never reported without quality, and adoption is never reported without retention.

Client identity withheld
August 2026 · Rev 1.0

Metric	Owner	Cadence	Target	Live from
Tier 1 — Enterprise outcome · CEO and PMC · quarterly				
Vitality index Revenue from products launched in the last three years	CFO / Category Directors	Quarterly	20%	**Now**
Market share by category Unit and revenue share per product category, by region	Product Category Directors	Quarterly	Defend share in the two lead categories	**Now**
Connected attach rate Products activated and paired within 30 days of registration	Digital lead	Monthly	Set at Release 1	**R1 launch**
Recurring revenue ARR, with paid conversion and net revenue retention	Digital lead	Monthly	Against the five-year ramp	**First revenue**
Tier 2 — Portfolio and program health · PMC · monthly				
Gap to target, per release Bottom-up finish date against committed release date, R1 / R2 / R3	Program Manager	Weekly	Zero, or escalated	**60 days**
Decision latency Days from decision requested to decision recorded, by deciding body	EPMO	Monthly	PMC under 14 days; Triad under 3	**60 days**
Critical-path role fill rate Core roles named and allocated on every committed team	Program Manager	Monthly	100% on committed programs	**60 days**
Business case forecast accuracy Actual 12-month revenue against the forecast approved at the DP gate	EPMO	Quarterly	Within 20%	**60 days**
Quality and economics · Product Owners, Quality and Finance				
Field quality Rolling twelve-month RMA rate by SKU, test-in-spec shown separately	Quality lead	Monthly	At or below current line rates	**Now**
Ecosystem quality Pairing, connection, data sync, OTA update and app crash rates	Software Technical Owner	Weekly	Set at Release 1	**R1 launch**
Cycle time Gestation, development and ramp, measured as three separate segments	EPMO	At closure	Reduction against class baseline	**60 days**
Time to break-even Months to positive contribution; CAC payback for software	Product Owner / Finance	Post-launch	Against the DP business case	**Now**

Live from
Now — the data already exists. No new build.
60 days — stand up within two months. No new systems, no new headcount.
R1 launch — needs telemetry specified as a Release 1 requirement. Cannot be retrofitted.
First revenue — needs subscription billing in place.

Deliberately not on this wall
The five team operating metrics. Schedule movement, float consumed, aged decisions, requirements volatility and team dedication are the Leadership Triads' weekly pack. They feed gap to target and decision latency above.
Twenty supporting metrics. Owned-IP share, platform reuse, portfolio load, kill rate, warranty cost, cost of delay and the rest stay in the specification, reviewed at the gate rather than on the wall.

Figure 7. The one-page dashboard, reproduced from the engagement with client identifiers removed. Twelve metrics, three tiers, 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 six-row weekly scorecard deserves its own note. Its two rows that refresh weekly, gap to target with milestone trend and critical-path talent coverage, 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 twelve 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 twelve real, the build sequence, and the four decisions a leadership team has to take in the room.

Six definitions to settle before measurement starts

#	Decision required	Recommendation
1	Where the cycle-time clock starts	First documented idea discussion, not team formation. Report gestation as its own segment so the front end cannot hide.
2	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.
3	What counts as a product for the vitality index	A distinct product family launch, not a SKU. Derivatives and variants do not reset the clock.
4	Break-even for software	Customer-acquisition-cost payback and months to recover, reported separately. Do not force the hardware definition.
5	Whether test-in-spec returns count as quality failures	Track separately, report both. Real cost and real dissatisfaction, but the corrective action is usability, not reliability.
6	What “connected” means for attach rate	Activated and paired within 30 days, with a separate 90-day active figure. Sold is not connected, and registered is not used.

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

The second item is the constraint on the whole schedule side of the framework. 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 percentage before that exists produces a number each function can dispute on its own terms, which is worse than publishing nothing. The sixth item carries commercial consequences beyond measurement: the definition of attach rate will shape sales compensation, channel behavior, and what the platform business case is allowed to assume. Set it loosely and the revenue forecast is built on customers who registered once and never returned.

Sequence the build around what exists

The build sequence respects data reality rather than framework elegance. This quarter: vitality index, market share, field quality and time to break-even, all from data already held in finance, market research and the returns database. Immediately, as a blocker: reconcile the plan of record, owned by the EPMO with the portfolio committee arbitrating scope disputes, on a thirty-day clock; per-program gaps publish in the interim and only the portfolio roll-up waits. Within sixty days: gap to target, decision latency, role fill rate, forecast accuracy and cycle time, with no new systems and no new headcount. The decision log behind decision latency is deliberately minimal: a four-field register kept by the EPMO from existing gate and committee minutes (decision, date requested, deciding body, date recorded); a spreadsheet, not a system. Before the first connected release locks its architecture: ecosystem telemetry specified as a product requirement. At launch and after: attach rate, then recurring revenue at first subscription billing. For the sixty-day five, honesty about meaning matters as much as speed: forecast accuracy and cycle time publish definitions and start accumulating in sixty days, and produce their first trustworthy readings only as launches close. Publishing the easy four immediately is not cosmetic; it establishes the review rhythm before the harder metrics arrive [28].

Four decisions for the room

Approve the twelve and name an owner for every one. A number without a named owner is a slide. Several owners in the framework are roles that are unfilled or contested, so approving the set is also an organizational statement. Until a role is filled, the tier's governing body holds the number, and the fill date appears on the dashboard itself.

Commit to one reconciled baseline before any schedule metric is published. Accept that gap to target stays dark until the plan of record exists, rather than publishing a contested number in the interim.

Fund ecosystem telemetry as a first-release requirement. An architecture decision with a closing window, owned by the software team rather than by a future analytics project.

Fix the size of the set and adopt the displacement rule. Twelve is the ceiling. Additions happen at the annual review by displacing an existing metric, with the reason recorded.

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 Tier 1 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 twelve 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 twelve numbers that say whether new products are moving, whether they are good, and whether they make money, 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, January 2019. https://lateralworks.com/ideas/integrated-core-team

[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, January 2019. https://lateralworks.com/ideas/targets-and-trends

[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, January 2019. https://lateralworks.com/ideas/manage-the-fuzzy-front-end

[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