THE PLAYBOOK

The Delivery

Running an agency so the work ships without heroics. Utilization, scope, handoffs, and the conversations that protect margin.

Paul Prado Pacardo

Executive Assistant & Operations · me.kabuhayan.app

Contents

OPENING

Agencies rarely die from lack of clients

Most struggling agencies I have seen were not short of work. They were short of a way to do the work that did not depend on particular people staying late.

The pattern is consistent enough to be almost boring. Sales goes well, delivery is heroic, margin is thinner than anyone expected, and nobody can explain precisely where it went. The month with the most revenue was somehow the most stressful and not obviously the most profitable.

That is an operations problem wearing a growth costume, and it is fixable with a handful of unglamorous mechanisms: a gate on what comes in, scopes with edges, a delivery rhythm, and the willingness to have one uncomfortable conversation before it becomes an expensive one.

Nothing here is clever. All of it is skipped by agencies under pressure, which is precisely when it matters most.

Written for owners and for the operations person who has been quietly holding delivery together without the authority to change how it works.

CHAPTER 1

The two numbers that decide everything

Agency profitability rests on two ratios, and they multiply rather than add, which is why small movements in each produce large movements in the result.

Utilization is the share of available hours spent on client work. Not hours worked. Hours available. A team that is technically busy but spends a third of its week on internal churn has a utilization problem regardless of how tired everybody is.

Realization is the share of client hours you actually get paid for. Work outside scope, rework after unclear feedback, hours written off to keep a relationship warm. This is where most agencies lose money invisibly, because those hours were worked, recorded and then quietly discounted.

Utilization	×	Realization	=	What you actually earn
hours on client work		hours actually billed		from the hours you paid for

Eighty percent of eighty percent is sixty-four. Two respectable numbers, one poor result.

Figure: the two ratios multiply, which is why both must be watched

Track both weekly, per client, against contracted hours. Most agencies track neither and discover the problem at the quarter end, by which point the hours are spent and the conversation is retrospective.

CHAPTER 2

Where agency margin actually leaks

Four leaks, in roughly the order of how much they cost and how invisible they are.

01 Scope that grew without a conversation

The extra round, the small additional deliverable, the favour. Individually trivial, collectively the largest single leak in most agencies, and entirely undocumented because nobody wanted to be difficult.

02 Rework caused by unclear feedback

Work redone because approval was vague or because the wrong person approved it. These hours are worked twice and billed once, and they are usually recorded as the team being slow.

03 Senior people doing junior work

The most expensive hours in the business spent on things that did not require them, usually because delegating would have taken longer this once. This is the leak the owner personally causes most often.

04 Internal work nobody counts

Status meetings, rebuilding the same report, chasing information. It never appears against a client and it consumes the utilization half of the equation entirely.

The counterintuitive part, and it holds up across the agencies I have seen: the most profitable owners bill less of their own time, not more. An owner heavily loaded with delivery cannot sell, cannot fix the system, and becomes the constraint on everything at once.

CHAPTER 3

The intake gate

Most delivery problems are sales problems that arrived a month earlier. A gate between winning work and starting it is the cheapest possible intervention.

Nothing starts without a written scope with an edge. Not a proposal describing ambition. A document stating what is included, what is explicitly not, how many rounds of revision, and what happens when something outside it is requested. The exclusions matter more than the inclusions.

Nothing starts without capacity confirmed. Check the actual availability of the actual people before a start date is promised. An agency that sells a January start it cannot staff has created a delivery crisis and called it a good month.

Nothing starts without one named client-side approver. Work approved by committee is work revised indefinitely. One name, agreed in writing, and everything else is input rather than instruction.

Nothing starts without a kick-off that states the rules. How feedback is given, in what form, by when, and what happens when a deadline slips on their side. Saying this warmly at the start is easy. Saying it in month two is a confrontation.

THE GATE IS NOT BUREAUCRACY

It is four checks and takes about twenty minutes. Every one of them prevents a specific conversation that would otherwise happen later, under pressure, with money attached, and with somebody already disappointed.

CHAPTER 4

Scoping so the work has an edge

An unscoped project does not stay unscoped. It gets scoped later, informally, by whoever is least comfortable saying no, which is usually the person doing the work.

Define done for each deliverable, not for the project. Project-level definitions are too abstract to enforce. Deliverable-level ones can be pointed at, which is what you need when a request arrives that sits just outside.

State the number of revision rounds and what a round is. Two rounds means consolidated feedback twice, not two months of individual comments. Say that explicitly, because the ambiguity is where realization goes to die.

Price the change, always, even at zero. When something out of scope is requested, respond with what it costs in time or money before doing it. You may well decide to absorb it, and you should say that you are absorbing it. Free work that is invisible teaches the client that the boundary is imaginary; free work that is named builds goodwill and holds the line at once.

Attach a dependency clause. Most late projects are late because the client was late, and without a written expectation the delay silently becomes yours, along with the overtime.

> **The scope document is not a legal instrument. It is the thing that lets a junior person say no without needing courage.**

CHAPTER 5

The delivery rhythm

Agencies run on a weekly cycle whether or not anyone designs one. Designing it costs an hour and removes most status meetings.

Monday, load and check. Who is working on what this week, against contracted hours. This is where over-allocation is caught, while it is still a scheduling problem rather than a Thursday night.

Midweek, unblock only. A short session for things that are stuck, not a status round. If nothing is stuck it does not happen, and cancelling it is a success rather than a failure.

Friday, burn and forecast. Hours used against hours sold per client, and anything trending toward overrun flagged now. **An overrun spotted at forty percent through is a conversation. The same overrun spotted at ninety percent is a write-off.**

Monthly, realization review. What was written off and why, per client. Patterns appear quickly and they are almost always concentrated: one client, one deliverable type, or one approver.

THE REPORT THAT CHANGES BEHAVIOUR

A single weekly line per client: hours sold, hours used, percentage through the project. Circulate it to everyone who touches delivery. Visibility does most of the work that enforcement would otherwise have to do.

CHAPTER 6

Handoffs between people

Agency work moves between people constantly, and every transfer is a chance to lose context, repeat work, or drop something quietly.

Sales to delivery is the expensive one. The person who sold it knows what was promised, including the things said warmly in a meeting that never reached the scope document. If that transfer is a forwarded email, the delivery team will discover the promises one disappointment at a time.

Make it a fifteen-minute conversation with a written summary: what was sold, what was implied, who the client actually is, what they are anxious about, and anything promised that is not in the scope. That last item is the whole point of the meeting.

Between delivery people, hand over the state, not the file. Where it is, what was decided, what was tried and rejected, and what the next person should not redo. Rejected approaches are the most valuable and least transferred piece of context in creative work.

Into support or retainer, transfer the relationship deliberately. Introduce the new person while the old one is still present. A client silently reassigned to somebody they have never met is a client already partway out the door.

Every handoff should produce a written artefact, however short. Verbal handoffs feel faster and are the origin of most of the rework that appears later as a capability problem.

CHAPTER 7

The uncomfortable client conversation

Everything above exists to make one conversation happen earlier, because it is the conversation that decides whether a project is profitable.

It is the one where you say that what is now being asked for is beyond what was agreed, and here is what it would take.

Have it at forty percent, not ninety. Early, it is a scoping discussion between professionals. Late, it is a complaint about a bill, and by then you have already done the work and lost the leverage entirely.

Lead with the goal, then the constraint. "You want this live before the launch. What has been added means the current plan does not get there. Here are two ways it can." You are solving their problem, not defending your invoice.

Give options with consequences, not a refusal. Extend the timeline, extend the budget, or reduce the scope. Three doors, all of them open, and the client chooses. A flat no is a relationship problem; three doors is a conversation.

Then write it down the same day. A verbal agreement about scope is not an agreement about scope, and in six weeks two people will remember it differently and both will be sincere.

The agencies that hold margin are not the ones with tougher contracts. They are the ones that have this conversation early, kindly, and every single time.

CLOSING

Boring is the goal

An agency that runs well is unremarkable from the inside. Work starts when it was supposed to, people are allocated before Monday, overruns are noticed at forty percent, and the difficult conversation happens while it is still small.

None of that is a growth strategy and all of it is why growth survives contact with delivery.

If you change one thing, make it the weekly line per client showing hours sold against hours used. It costs almost nothing, it exposes the leak that does the most damage, and it forces the early conversation by making the late one impossible to be surprised by.

If you change two, add the intake gate. Four checks, twenty minutes, before anything starts.

Most agency chaos is not caused by demanding clients or unreliable people. It is caused by work beginning before anyone decided precisely what it was, and by nobody looking at the numbers until the month had already happened.

Paul Prado Pacardo

Senior Executive Assistant, Operations and Project Manager.
Ten years supporting C-level leadership. Available for remote work.

me.kabuhayan.app
linkedin.com/in/paulpacardo
pacardopaul18@gmail.com