ValueLogics.ai

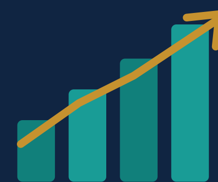
THE ROI AGENT COMPANY

Sarnosites

FRAMEWORK · THE EVIDENCE LAYER ABOVE CMMI-ACQ AND P3M3

Value Buying Maturity

Process maturity asks whether you do the thing. Evidence maturity asks what it produced. Five levels, graded on what your claims would survive.



FRAMEWORK PAPER

5	**12**	**2**
LEVELS, GRADED ON EVIDENCE	DIAGNOSTIC QUESTIONS	LEVELS THE BUYER TRAILS THE SELLER

> **What this grades.** Not your buying process — other models already do that, and do it well. This grades **the evidence your buying process produces**: whether the claims made to justify what you bought would survive being audited. It is the layer that sits above CMMI-ACQ and P3M3, not a replacement for either.

Why another maturity model

There are good buyer-side maturity models already. **CMMI-ACQ** grades acquisition process discipline — solicitation, supplier agreements, requirements. **P3M3** runs seven perspectives and one of them is Benefits Management, five levels, optimising at the top. **PMI's Benefits Realization Management framework** covers identify, execute, sustain. None of this is missing.

What all of them grade is *process*. Do you have an approach. Is it documented. Is it applied consistently. Is it improved over time. Those are the right questions about a process and they are the wrong questions about a claim.

THE DISTINCTION THIS MODEL TURNS ON

Process maturity asks whether you do the thing. Evidence maturity asks what it produced.

These are different axes, and the gap between them is the oldest known failure mode of every process maturity model ever written: an organisation can become reliably mature at doing a thing badly.

You can sit at P3M3 Level 4 on Benefits Management — measured, controlled, systematically managed — and still have sixty per cent of your claim lines fail on evidence. Conformance says the review happened. It says nothing about whether anything in the minutes would survive an auditor.

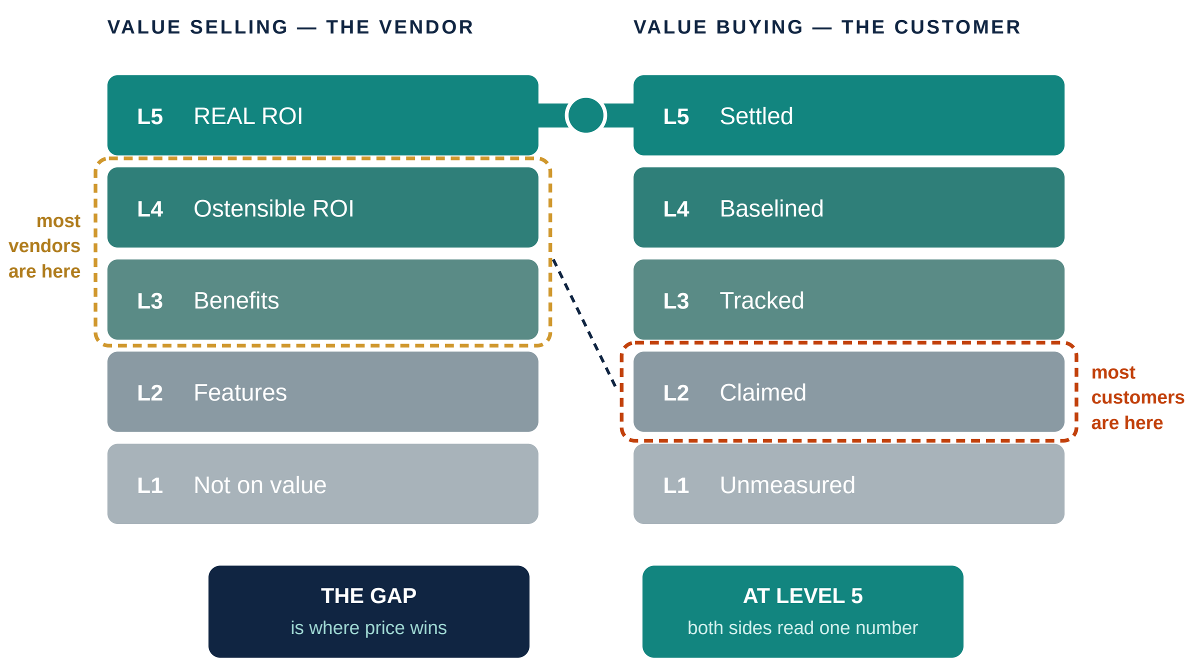
Where it sits

MODEL	WHAT IT GRADES	THE QUESTION IT ASKS
CMMI-ACQ	The buying process	Do you acquire in a disciplined way?
P3M3 — Benefits	The benefits process	Do you manage benefits repeatably?
Value Buying	The evidence produced	Would your claims survive an audit?

> **A lineage worth claiming.** CMMI levels 4 and 5 brought statistical process control to software — measure variance continuously, correct it while it is still variance rather than outcome. That discipline was never pointed at the business case. Continuous measurement against design intent is the same idea applied to value instead of defects, and there is no reason the business case should be the one artifact exempt from it.

The two ladders

ValueLogics has graded the seller for twenty-two years: the Value Selling Maturity Model runs from not selling on value at all, through features and benefits, to ostensible ROI, to real ROI. Set the buyer's ladder beside it and something uncomfortable becomes visible.



Commoditisation is usually diagnosed as a seller problem. The ladders say otherwise — the buyer is two levels behind, and price is the only dimension they are equipped to judge. The two ladders are separate everywhere except the top.

The vendor is usually two levels ahead of the customer. That inversion explains something the Paradox of Excellence gets almost right.

Commoditisation is normally diagnosed as the customer losing interest — excellence becoming invisible, attention drifting, the remedy therefore being communication. The ladders suggest a harder cause: the customer has no instrument. Price is not what they care about most. It is the only dimension they are equipped to judge.

You cannot fix a buyer's missing instrument with a seller's better communication. That is why the remedy has been prescribed for twenty years and the paradox is still here.

The five levels

Each level is defined by an artifact that either exists or does not. No level is awarded for intent, effort, or having a policy that says you would.

L1 — Unmeasured	No business case is required, or one is required and no copy survives approval. Purchases are justified by preference, urgency or price.
L2 — Claimed	A business case exists at approval. Its numbers are vendor-supplied or internally estimated, unverified. Nobody owns the document after signature.
L3 — Tracked	The case has a named owner after signature. Adoption and activity are reported at intervals. Nothing is compared to a recorded pre-state, and nothing is stated in money finance recognises.
L4 — Baselined	A dated pre-state exists, recorded before the work began. Matched actuals are measured on the same definition. Attribution is asserted but not tested against alternatives.

L5 — Settled	Every claim line carries source, method, a tested attribution and a grade. Results are published regardless of sign. Drift is watched in term against design intent, and the next decision is priced against the measured gap.

What each level costs

L1	You cannot tell a good vendor from a cheap one, so you select on price and call the result value for money.
L2	You pay for benefits you have already discounted to zero. The vendor prices the benefit in; you refuse to credit it. Both sides lose and the deal closes on rate.
L3	Reporting creates the appearance of governance with none of the evidence. Reviews are well attended and decide nothing.
L4	You can prove the outcome but not the cause, so your strongest results remain challengeable by anyone who wants to challenge them.
L5	Renewal stops being a negotiation and becomes a calculation.

The diagnostic

Twelve questions, four blocks of three. Every answer is yes or no, and every yes has to be provable by producing the artifact — not by recalling that it exists somewhere.

Scoring. Your level is the highest block in which all three answers are yes, and all lower blocks are also all yes. A single no caps you at the level below. There is no partial credit, because there is no partial evidence: a claim line either carries its source or it does not.

Level 2 — Claimed

- For your last three major software purchases, can you produce the business case that justified each one?
- Does each state a quantified benefit, not only a cost?
- Does each carry a date and a named author?

Level 3 — Tracked

- Is a named person accountable for each case *after* signature?
- Has any of the three been formally reviewed since go-live?
- Does a report exist comparing post-go-live performance to what the case said?

Level 4 — Baselined

- For at least one, is there a dated pre-state measurement taken *before* the work started?
- Is there a post-state measured on the same definition, by the same method?
- Are the benefits stated in a unit your finance function already reports?

Level 5 — Settled

- For each claimed benefit, is a specific attribution stated — with an alternative explanation considered and ruled out or discounted?

- ▶ Has any post-implementation review reported a result *below* the original claim, in writing, in the same format as a favourable one?
- ▶ Is value drift monitored during the term, rather than only at review points?

> **The question most organisations fail on.** It is the second one in Level 5 — has any review ever reported a result *below* the original claim, in writing, in the same format as a favourable one. Almost nobody can say yes. And until the answer is yes at least once, every favourable result the organisation has ever published is worth precisely what an unfalsifiable claim is worth, which is nothing. The willingness to print a shortfall is not a governance nicety. It is the thing that makes every other number mean something.

What the ladder cannot grade

One honest limit, stated before anyone finds it: this model grades the evidence a decision left behind. It does not grade whether the decision was worth making.

An organisation can reach Level 5 on a requirement that was never going to create value — frozen baseline, matched actuals, tested attribution, published regardless of sign, and a flawlessly evidenced record of something that did not matter. Perfect measurement of a worthless thing is still perfect measurement. It is also still worthless.

WHERE THE CHAIN ACTUALLY STARTS

A real business requirement
→ the dominant constraint
→ the measure
→ alignment
→ proof

> ROI is not a calculation performed on a project. It is the link between a requirement that creates value when addressed and the evidence that it did. Break the first link and everything downstream is arithmetic about nothing.

This matters because it is a third, separate failure mode, and it is the most expensive of the three. A claim can fail because the value was created and the evidence erased. It can fail because the measure was a proxy standing at a distance from the problem. Or it can fail because **the requirement was delivered perfectly and was never value-bearing to begin with**. Only the third one is invisible to a well-run programme, because every status report is green all the way to the end.

Establishing that a requirement is value-bearing, and finding the dominant constraint it sits against, happens before any of this. That is the Problem Pyramid's work, and no amount of evidentiary rigour downstream substitutes for it.

What moves you up

- **An ROI Replay places you.** Running the four proof tests across an existing claim set returns a distribution of grades, and that distribution is the level. It is a measurement, not an interview.
- **A frozen baseline lifts you to 4.** It is the single highest-leverage artifact in the model, and the only one that cannot be created retroactively past a certain point — the record decays with the people who remember it.
- **ROI DriftWatch holds 5.** Level 5 is not a state you reach, it is a state you maintain. Continuous measurement against design intent is what stops an

organisation sliding back to 3 the moment the programme that cared about it finishes.

Using it as a gate

The diagnostic is cheap to run and expensive to argue with, which makes it the right front door rather than a marketing instrument bolted on afterwards.

It qualifies in both directions. An organisation at Level 1 or 2 is not yet ready for a Replay — there may be nothing left to replay, and finding that out early is worth more to them than an engagement that discovers it slowly. An organisation at 3 or 4 has evidence sitting unused and a short window in which to recover it. And the score itself is the finding: it is a statement about the organisation's capacity to know whether anything it buys works, which holds for the next vendor and the one after that.

The honest framing for anyone who asks. This model does not replace CMMI-ACQ or P3M3 and does not compete with them. Those frameworks grade whether you run the process. This one grades what the process left behind. An organisation serious about both should expect to score lower here than there — and the distance between the two scores is the most useful number in this document.