SUPPLY CHAIN RESEARCH · EDITORIAL FEATURE

The Procurement Savings Illusion

Why the Number Procurement Reports Is Not the Number Finance Sees

Baseline construction, cost-avoidance accounting, and the absence of finance validation produce two defensible numbers describing the same year. The strongest finding in this research is an absence: no independent, sampled study of savings realization exists.

Independent, vendor-neutral research for supply chain and technology leaders

Supply Chain Research | supplychainresearch.com | 2026

Contents

Why the Number Procurement Reports Is Not the Number Finance Sees

01 Executive summary

Procurement functions report savings annually, the figure is large, and it is presented to boards as a contribution to earnings. Finance functions observe what was actually paid, and the two numbers rarely agree. The gap is not usually fraud and it is not usually incompetence. It arises from three structural features of how savings are constructed and reported: the baseline against which a saving is measured is chosen by the party that will report the result, cost avoidance is measured against a counterfactual that by definition never occurred, and in most organizations no finance function validates the claim before it is published.

The strongest finding in this research is an absence rather than a number. There is no independent, sampled, peer-reviewed study establishing what proportion of reported procurement savings reaches the financial statements. Every realization-rate figure in general circulation originates with either a consultancy that sells procurement advisory work or a source-to-pay software vendor that sells savings-tracking software. That is worth stating plainly rather than resolving by selecting the most credible-sounding interested estimate. The one truly disinterested evidence base is public-sector audit, where auditors have repeatedly found cost reduction and cost avoidance reported as a single blended figure, and which establishes that the conflation occurs without establishing how often. We say honestly that procurement creates substantial value and that a good deal of it is invisible to the profit and loss account by its nature. The argument here is that combining traceable and untraceable value into one headline number destroys the credibility of both, and that the remedy is separation and validation rather than a smaller claim.

30 to 40%	Zero	75 to 85%
the negotiated-to-realized gap described in practitioner literature, sourced entirely from interested parties	independent, sampled, peer-reviewed studies of savings realization located by this research	the delivery rate attributed to organizations with validated savings governance, again from an interested source

The argument in brief

- **Two defensible numbers describe the same year.** Procurement measures negotiated outcomes against constructed baselines. Finance measures what was paid. Neither is wrong and no reconciliation process exists in most organizations.
- **There is no neutral number.** Every realization figure in circulation comes from a consultancy selling advisory work or a vendor selling savings-tracking software. The absence of independent evidence is itself the finding.
- **The baseline decides the answer.** Historical price, budget, first quote, and should-cost produce different results from identical negotiating. The most gameable is the one solicited by the party reporting the outcome.
- **Three categories, one figure.** Hard savings, soft savings, and cost avoidance have different verifiability. Reporting them as one number costs the traceable portion its credibility.

- **The signature is the only test that settles it.** A saving the controller will sign for is a saving the organization can spend. One they will not sign for is a claim about a counterfactual, however well founded.

02 Two numbers, both defensible

The disagreement at the centre of this article is familiar to anyone who has sat in a review where a chief procurement officer and a chief financial officer discuss the same year and describe different outcomes. Figure 1 states it in the form the practitioner literature uses.

Figure 1. Two numbers, both defensible, describing the same year. Illustrative figures used throughout the practitioner literature. Neither function is wrong; they are measuring different quantities against different baselines over different horizons.

Procurement reports a figure built from negotiated outcomes. For each sourcing event, the team establishes what would have been paid absent the intervention, compares it with what will now be paid, multiplies by expected volume, and sums across the portfolio. Every step of that calculation is defensible on its own terms, the work behind it was real, and the resulting number is an accurate description of what the function achieved against the yardstick it applied. It is a measure of negotiating outcome.

Finance observes something different, which is the money that left the organization. The chief financial officer sees invoices paid and general ledger movements, and asks a narrower question: is the amount spent on this category lower than it was, and has the budget been reduced accordingly. That question is also legitimate, it is the one the organization's external reporting depends on, and it produces a smaller answer because it captures only the portion of procurement's achievement that manifested as reduced outflow within the period.

The two functions are therefore not in disagreement about facts. They are measuring different quantities and reporting both under the single word savings, which is a definitional failure rather than an analytical one. What makes it persistent is that no process exists in most organizations to convert one measure into the other. The reconciliation is possible, requires the two functions to agree in advance which terms will be held constant, and is performed by a minority of organizations. Where it does not happen, the annual conversation becomes an argument about credibility that neither party can win because they are not describing the same thing.

One clarification of scope belongs here, because this argument shares a family resemblance with a different one. There is a well-documented pattern by which performance metrics become targets and then become gamed, and it applies to forecast accuracy, service levels, and much else. The savings case is a specific instance with distinctive features: the counterparty is finance rather than another operating function, the disputed quantity has a formal accounting counterpart that either exists or does not, and the resolution mechanism is a validation signature rather than a better measurement definition. The rest of this article stays on those distinctive features.

There is an organizational feature that keeps the two numbers apart even when both functions want them reconciled, which is that they are produced on different calendars by different systems. Procurement records a saving when a contract is signed, in a sourcing or contract management system, attributed to the category team that delivered it. Finance records spending when an invoice is paid, in the general ledger, attributed to the cost centre that consumed it. The two records share no common key, and building one is a data exercise that neither function owns.

That absence of a common key is worth naming because it explains why the reconciliation described later in this article is harder in practice than it sounds in principle. Linking a negotiated saving to the ledger movements it should have produced requires mapping contracts to categories to cost centres to accounts, and in most organizations that mapping is incomplete. Establishing it is the unglamorous precondition for everything else, and it is generally the reason a reconciliation initiative stalls in its first quarter.

03 The cascade, as the field describes itself

The practitioner literature has a standard account of how a negotiated saving decays on its way to the financial statements, and it is worth setting out both because it is broadly credible and because of who publishes it. Figure 2 presents the model.

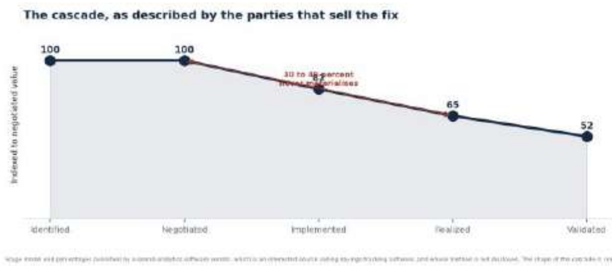

Figure 2. The cascade, as described by the parties that sell the fix. The stage model and percentages are published by a spend-analytics software vendor whose method is not disclosed; the shape is corroborated by practitioner accounts, the specific figures are not independently verified.

The model distinguishes four stages after identification. A saving is negotiated when the contract is signed at the improved rate. It is implemented when the new rate is loaded into the purchasing systems and the organization is actually able to transact at it. It is realized when purchases are made at that rate and the lower amount is paid. It is validated when finance confirms that the reduction is visible and reflected in the budget or reforecast. Value is lost at every transition, and the published figures put the negotiated-to-realized gap at thirty to forty percent, the realized-to-validated gap at a further fifteen to twenty percent, and the delivery rate for organizations with mature validation governance at seventy-five to eighty-five percent.

The mechanisms behind each transition are unglamorous and well understood by anyone who has run a category. Between negotiated and implemented, the new pricing may not be loaded, the catalogue may not be updated, and the requisitioners may not know the contract exists. Between implemented and realized, the organization may not buy the volumes assumed, may buy a different mix, or may continue buying from the incumbent supplier out of habit. Between realized and validated, the reduction may be real and invisible because other movements in the same account swamped it, or because no one reduced the budget so the money was simply spent elsewhere.

The provenance of these figures requires attention and is the subject of the next section. The stage model itself is published by a spend-analytics vendor whose product exists to track savings through exactly these transitions, which is to say the party describing the problem also sells the instrument for measuring it. This does not make the model wrong; it is a sensible decomposition that matches practitioner experience across many organizations. It does mean the specific percentages should be treated as the field's description of itself rather than as measurement, and this article uses them on that basis and says so wherever they appear.

One transition in the cascade deserves separate attention because organizations consistently underestimate it, which is the step between implemented and realized. A contract loaded correctly into the purchasing systems delivers nothing unless the people who buy actually use it, and the proportion of spend that flows through negotiated agreements varies enormously between organizations and between categories within the same organization. Indirect categories with many requisitioners and low individual transaction values are where the leakage concentrates.

This matters because it is the one stage where the remedy is entirely within the organization's control and requires no supplier cooperation. Enforcing catalogue use, restricting purchasing channels, and requiring approval for off-contract transactions recover savings already negotiated at no additional commercial cost. Functions that measure contract compliance by category alongside their savings figure generally find that a modest improvement in compliance delivers more realized value than a further round of negotiating, and it is considerably cheaper to obtain.

04 There is no neutral number

This research set out to establish what proportion of reported procurement savings actually reaches the financial statements. It did not find an answer, and the reason it did not is the most useful finding available. Figure 3 sets out the position.

Figure 3. There is no neutral number. Every realization-rate figure in general circulation originates with a party selling procurement advisory services or savings-tracking software. No independent, sampled, peer-reviewed study of realization was located.

The figures in circulation divide into two groups by source. The stage-cascade percentages, the negotiated-to-realized gap and the mature-organization delivery rate, come from source-to-pay and spend-analytics software vendors whose products track savings through the stages being described. The benchmark figures about world-class procurement performance, the proportion of spend influenced, the multiple of savings delivered relative to peers, and the return on investment ratios, come from a benchmarking consultancy that sells both the benchmark and the advisory work to close the gap it identifies. Both groups have a direct commercial interest in the finding that reported savings substantially exceed realized savings, because that finding is the case for their product.

This does not mean the figures are wrong. Vendors and consultancies frequently have the best available view of a phenomenon precisely because they observe many organizations, and a party with an interest in a conclusion may still report it accurately. It means that the figures are unverifiable: the samples are not disclosed, the methods are not published, the definitions of realized and validated are the vendor's own, and no independent party has attempted to replicate any of it. A figure with those properties can be quoted as a claim and cannot be relied upon as a measurement.

The academic position is worth stating precisely because its absence is surprising. There is an established peer-reviewed venue for research in this field, publishing work on purchasing and supply management including savings measurement and value leakage. This research located no canonical sampled study within it producing a realization rate that the field cites. That is unusual for a question of this commercial importance, and the most likely explanation is practical rather than sinister: the data required would be commercially sensitive, would need to be extracted from both procurement and finance systems in the same organizations, and would require a definition of realization that the participating firms agreed in advance. Those are hard conditions to satisfy across a sample large enough to generalise.

The practical consequence for an organization is that it cannot benchmark its own realization rate against an industry figure, because no reliable industry figure exists. It can, however, measure its own, and that measurement is more useful than any benchmark would be. An organization that reconciles one year of reported savings against its own ledger has a number specific to its own definitions, its own categories, and its own accounting, and it can track that number over time. Section twelve sets out how to do this. The absence of an external figure is an argument for internal measurement rather than an obstacle to it.

A related observation concerns what the field cites when pressed for evidence, which is usually another secondary source rather than a study. Tracing any widely-quoted realization figure backward tends to produce a chain of practitioner articles, each citing the last, terminating at a vendor publication with no disclosed method. This is a familiar pattern in commercial research literature and it is worth performing the exercise once for any figure an organization intends to rely upon, because the chain is generally shorter than expected and its origin is generally interested.

The practical discipline that follows is to require a primary source before any external figure enters a business case or a board paper. Where the primary source turns out to be a vendor, the figure can still be used with the attribution attached, which costs one clause and prevents a marketing estimate from acquiring the standing of a finding. Where no primary source can be located at all, the figure should be dropped rather than softened with a qualifier, because a number nobody can trace is not evidence in any useful sense.

05 The baseline decides the answer

Of the three structural features named in the summary, baseline construction is the one that does most of the work, because the baseline determines the reported outcome before any negotiating has occurred. Figure 4 compares the common choices.

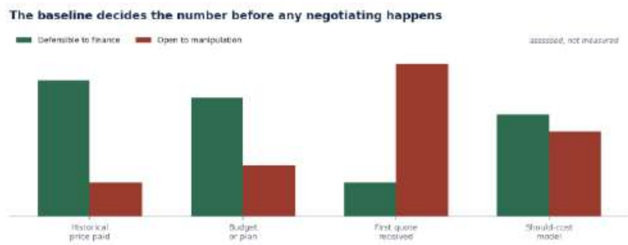

Figure 4. The baseline decides the number before any negotiating happens. Qualitative assessment rather than measured data. First quote received is the most gameable, because the quote against which the saving is measured is solicited by the party that will report the saving.

Historical price paid is the most defensible baseline and the most conservative. It is verifiable in the ledger, it is difficult to manipulate, and finance recognises it immediately because it is the same number finance has. Its weakness is that it ignores market movement: in a falling market, measuring against last year's price credits procurement with a decline it did not cause, and in a rising market it understates a genuine achievement in holding an increase below the index. Organizations using this baseline should adjust for the relevant index and should say whether they have.

Budget or plan is the baseline finance finds most natural, because it is the comparison the organization already runs. Its weakness is that a budget is itself a negotiated number, frequently set with an allowance for the savings that are expected, which makes measuring savings against it partly circular. Where the budget was set on the assumption of a five percent reduction, delivering a five percent reduction produces zero measured saving against budget and a substantial saving against last year's price, and both statements are true.

First quote received is the most gameable of the four and it is common. The saving is measured as the difference between the initial quotation and the final agreed price, which means the number depends entirely on the initial quotation, which was solicited by the team that will report the result and which the supplier had every reason to pitch high. No manipulation is required for this to inflate: a supplier opening at an ambitious number and settling at a normal one produces a large reported saving and no change whatsoever in what the organization pays. Should-cost modelling is analytically the strongest approach and depends on assumptions about input costs, yields, and margins that the same team selects, which makes it defensible where the model is documented and reviewed and unverifiable where it is not.

The single most valuable governance change available in this area is to fix the baseline before the sourcing event rather than when the result is reported. A baseline agreed in advance with a finance counterpart, documented, and applied regardless of outcome removes the largest source of dispute and costs nothing. Where the baseline is chosen after the negotiating is complete, the choice will be made by someone who now knows which baseline produces the better number, and no amount of integrity fully neutralises that position.

A fifth baseline appears in some organizations and deserves a caution, which is the market benchmark supplied by a third party. Measuring a negotiated price against an index of what comparable organizations reportedly pay has intuitive appeal and introduces two problems. The comparison population is rarely disclosed in enough detail to establish comparability of specification, volume, and service level, and the party supplying the benchmark frequently also sells advisory work premised on the gap it identifies.

Used carefully, a market benchmark is a useful sense check on whether a negotiated outcome is competitive. Used as the baseline from which savings are computed, it imports an external and unverifiable number into the middle of the organization's own reporting, and it will not survive a question from an auditor about where the benchmark came from. The safer arrangement is to use the benchmark diagnostically, to decide where to focus effort, and to compute the reported saving against an internal baseline that finance can verify.

06 Three categories reported as one

The second structural feature concerns what is being counted, and the problem is that three different things with three different verifiability properties are routinely combined into a single headline. Figure 5 separates them.

Figure 5. Three different things, one reported figure. All three represent real work and two represent real value that never appears in the financial statements. Combining categories with different verifiability destroys the credibility of the whole number.

Hard savings reduce a recurring budget line. The organization was paying a certain amount for a certain thing, it now pays less, the invoice is smaller, and the budget can be cut by the difference. This category is fully traceable, finance can see it, and it is the only category that unambiguously converts into spending capacity elsewhere. It is also usually the smallest of the three in reported terms, which is part of why the categories get combined.

Soft savings capture productivity or process improvement without a corresponding reduction in any budget line. A process that took four hours now takes one; the organization has gained capacity, nobody was released, and no line item fell. The value is real in the sense that the freed capacity can be applied elsewhere, and it is invisible to finance because no money stopped leaving. Whether it becomes financial value depends entirely on whether the organization subsequently reduces headcount or grows output without adding cost, which is a management decision taken separately and frequently not taken at all.

Cost avoidance is the most contested and the most interesting. A supplier proposed an increase of eight percent, the negotiation settled at three, and the organization records the five-point difference as a saving. Nothing in the ledger falls; the amount paid rises. The claim is that spending is lower than it would have been in a counterfactual world, and the strength of the claim depends entirely on how credible that counterfactual is. Where an index, a published tariff, or a comparable market movement supports it, the claim is well founded. Where the counterfactual is the supplier's opening position, it is a claim about a negotiating posture. One vendor operating in this market states the limitation with unusual candour, observing that cost avoidance can be overstated through hypothetical claims that are commercially plausible but not demonstrably attributable, and that the main limitation is evidentiary.

The damage from combining these is asymmetric and worth being precise about. When a single figure blends fully traceable reductions with counterfactual claims, finance discounts the entire number rather than the untraceable portion, because it has no way to separate them. The hard savings, which were real and verifiable, lose the credibility they had earned. Procurement functions that separate the categories generally find that their hard-savings number, though smaller, is accepted without argument and can be relied upon in planning, which is worth considerably more than a larger contested figure.

A boundary case worth addressing is the multi-year contract, where the categorisation becomes ambiguous in a way that produces some of the largest disputed numbers. A five-year agreement negotiated at improved rates is frequently reported in full in the year it is signed, which attributes five years of benefit to a single period and inflates that year's figure by a factor the organization then has to sustain in the next one. The alternative treatment, recognising the benefit as it is consumed, is more conservative and matches how finance will see it.

The choice between these treatments should be made once and applied consistently, and organizations should be aware that the aggressive treatment creates a ratchet. A function that books multi-year value up front must find progressively larger deals to show growth, which distorts sourcing priorities toward large consolidations regardless of whether they are the right commercial move. Recognising benefit as consumed removes that pressure and produces a figure that finance can reconcile without adjustment.

07 Six reasons a real saving vanishes

Even where a saving is unambiguously hard, correctly baselined, and candidly reported, it may not be visible in the accounts, and understanding why is necessary before any reconciliation can work. Figure 6 lists the mechanisms.

Figure 6. Six reasons a real saving does not reach the ledger. Each can break traceability between a negotiated outcome and a general ledger line without anyone having done anything wrong.

Volume is the most common. A negotiated unit price reduction of ten percent, applied to volumes that grew by thirty percent, produces higher total spend in the category. The saving is real, the organization paid less per unit than it otherwise would have, and the account balance rose. Finance looking at the account sees an increase and procurement looking at the unit price sees an improvement, and both are describing the same transactions.

Mix operates similarly and is harder to see. The basket purchased in the current period is not the basket in the baseline, so a comparison of totals compares different things. Where the organization shifted toward higher-specification items for reasons unrelated to procurement, the category cost rises regardless of price performance. Adjusting for mix requires line-level data that both functions accept, and is the single most common technical obstacle to reconciliation.

The remaining four are equally mundane. Commodity and freight index movements can swamp a negotiated delta entirely, particularly in categories with volatile inputs. Currency movements separate the negotiated improvement from the amount paid where sourcing and settlement are in different currencies. Timing slippage places the benefit in a period the budget did not anticipate, so it appears as a favourable variance nobody attributes to procurement. And off-contract spend means the negotiated rate applies to purchases that were never made through the contract, which is the failure mode where the whole saving is theoretical because the organization continued buying elsewhere. Each of these is adjustable, and adjusting for them is what a reconciliation consists of.

A seventh mechanism belongs alongside these six and is the most consequential in organizations undergoing change, which is that the budget was never reduced. Where a saving is real, traceable, and correctly adjusted, and no one lowered the corresponding budget line, the money remains available and is generally spent. The category owner has a budget, the budget was not cut, and the freed amount is applied to something else that was previously unaffordable. The saving occurred and the organization is no better off in aggregate.

This is not a measurement failure and it cannot be fixed by better reconciliation, because the reconciliation will correctly show the saving. It is a governance decision about whether savings are harvested or retained, and different organizations legitimately choose differently. What matters is that the choice is made explicitly, because a function reporting savings into an environment where budgets are never adjusted is reporting a number that describes work performed rather than money released, and everyone should know which is being claimed.

08 What independent audit has actually found

If commercial sources are all interested, the question is whether any disinterested party has examined claimed procurement savings, and one class of examiner has. Figure 7 sets out what it does and does not provide.

Figure 7. The only disinterested examiner is a government auditor. Public-sector audit reports have repeatedly found cost reduction and cost avoidance reported as a single blended figure, and are not designed to produce a generalisable realization rate.

Government audit offices have no commercial interest in the conclusion, examine actual claims rather than survey opinions, and publish their findings. Reviews of defence acquisition programmes have found that savings and cost avoidance were reported together as a blended figure, which is precisely the conflation this article describes, occurring at the scale of national procurement. Work on contract audit has documented control weaknesses that permit unvalidated claims to persist. Broader reviews of fragmentation and duplication have recommended that agencies adopt metrics measuring cost reduction or avoidance, a formulation that itself illustrates how routinely the two are treated as one category.

What this evidence establishes is important and bounded. It establishes that the conflation is real, that it occurs in organizations subject to formal audit, and that it is a documented finding rather than a consultant's assertion. It does not establish a rate, because these audits examine specific programmes rather than sampling a population, and they were never designed to produce a generalisable figure. An organization cannot use them to estimate its own realization rate and can use them to establish that the underlying failure mode is well documented by parties with nothing to sell.

There is a gap in this evidence base worth flagging for readers who wish to go further. National audit offices outside the United States, and particularly those with an established practice of value-for-money reporting, produce examinations of claimed efficiency and procurement savings that would add an independent non-United States voice to this picture. This research did not complete a systematic review of that literature, and readers with a specific interest should consult those offices directly rather than relying on the summary here. The point of naming the gap is to avoid filling it with an interested estimate, which is the practice this article is arguing against.

The broader lesson from the public-sector material is about mechanism rather than magnitude. Where an external auditor with statutory access examines a savings claim, the conflation surfaces. Where no such examination occurs, it does not. That is an argument for internal validation of a specific and unusual kind: not review by another part of the same function, which shares the same definitions and incentives, but confirmation by a function with a different incentive and its own reporting obligations. In a commercial organization that function is finance, which is the subject of section ten.

There is one further reason public-sector material is worth reading even by commercial organizations, which is that the audit reports describe remediation as well as findings. Where an auditor has recommended that an agency distinguish cost reduction from cost avoidance, define its baselines, and validate claims through the finance function, the recommendation is public, has been tested against an actual programme, and carries no commercial interest. That is a better source for designing a governance change than a vendor's implementation guide.

The reports also illustrate what happens when the recommendations are not adopted, since audit offices revisit their prior findings. Repeated recommendations across successive reports indicate that the conflation is durable and does not resolve itself with better intentions, which is consistent with the incentive analysis in the next section. An organization can reasonably conclude that this failure mode requires a structural fix rather than an appeal to rigour.

09 The incentive structure underneath

None of the mechanisms described so far requires anyone to act in bad faith, and the article would be incomplete without examining why they persist even among people trying to report accurately.

Procurement functions are targeted on savings. The target is set annually, performance against it determines the function's standing and frequently individual compensation, and the number reported is the number by which the function is judged. That arrangement is entirely conventional and it creates a systematic directional pressure at every point where judgment enters the calculation: which baseline to apply, whether to include a category as avoidance, what volume assumption to use, and whether a partial implementation counts. Each individual judgment is defensible and the aggregate of many judgments made under directional pressure is a number biased upward.

Benchmark material reinforces the pressure. Published research from the principal benchmarking consultancy in this field consistently identifies cost savings as the leading priority reported by chief procurement officers, and offers comparisons in which world-class functions influence a substantially higher proportion of spend and deliver a multiple of the savings that typical peers achieve. Those figures are produced by a firm selling the benchmark and the advisory work, which is noted, and their effect on behaviour is real regardless of their accuracy: a function measured against them has an incentive to report on the most favourable defensible basis.

The corresponding incentive on the finance side is worth naming for balance. Finance has no particular interest in validating a savings claim, because doing so creates an expectation that the money is available and a corresponding obligation to reflect it in a reforecast, which constrains flexibility. A controller who declines to validate is not necessarily being rigorous; they may simply be avoiding a commitment. This is why the reconciliation described in section twelve needs to be a defined process with agreed rules rather than an appeal to finance for approval, which can be withheld for reasons unconnected to the merits.

The structural insight is that the reported figure is the output of a process in which one party has an incentive to maximise, the other has an incentive to withhold, and no agreed definition governs either. Fixing the definition and agreeing the rules in advance removes both incentives from the calculation, which is why the recommendations in this article concentrate on definition and process rather than on scrutiny of individual claims. Scrutinising claims after the fact is expensive, adversarial, and treats a systemic condition as a series of individual disputes.

The incentive question has a dimension that sits above the procurement function and is rarely discussed, which is what the organization does with the reported number externally. Where savings figures feed into investor communications, cost programme announcements, or transformation narratives, the pressure to sustain a headline is transmitted downward with considerably more force than any internal target carries. A function asked to support an externally communicated figure is being asked to produce a number rather than to measure one.

Organizations in that position should be particularly careful about the separation this article recommends, because the reputational exposure attaches to the combined figure. A hard-savings number validated by the controller can be communicated externally with confidence. A blended number that includes counterfactual claims cannot be defended if it is examined, and the examination that matters is not usually an internal one.

10 The signature test

There is a single test that settles most arguments in this area, and it is worth stating as the organising principle for everything that follows. Figure 8 frames it.

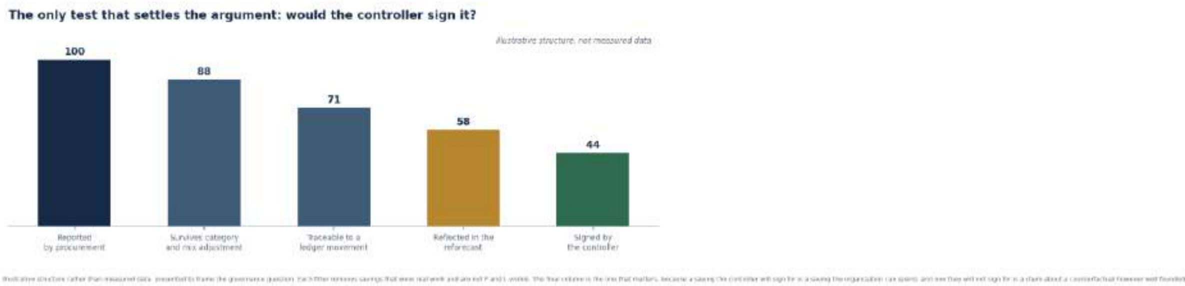

Figure 8. The only test that settles the argument: would the controller sign it? Illustrative structure rather than measured data. Each filter removes savings that were real work and are not visible in the profit and loss account.

The test is this: would the financial planning and analysis function or the controller sign a statement that this saving is real and is reflected in the reforecast. If the answer is yes, the saving is spendable, the organization can plan against it, and the argument is over. If the answer is no, the claim may still describe real value, and it is not a hard saving, and it should be reported in a different category with a different label. The test is binary, it is administered by a function with a different incentive, and it produces an answer that both parties can act on.

What makes this test powerful is that it is not an assessment of whether procurement did good work. A category manager who held a supplier to a three percent increase in a market moving at eight percent has performed extremely well, and the controller will decline to sign that the organization is spending less, because it is not. Both propositions are true simultaneously. Separating the question of performance from the question of financial visibility is what allows an organization to recognise the first without misrepresenting the second, and functions that adopt this framing generally find it improves their standing rather than diminishing it.

The test also has a diagnostic use during the year rather than at the end of it. A category team that asks the signature question at the point a contract is signed will discover immediately whether the saving is going to be visible, and if it is not, will learn what would have to change: a budget adjustment, a volume commitment, a catalogue update, or a compliance intervention to stop off-contract purchasing. Those actions are available in-year and are largely unavailable in the following March when the annual number is being contested. The test is therefore a management tool rather than an audit.

A practical caution belongs with this. The test requires a named finance counterpart with the authority to sign and the mandate to engage, and in many organizations no such person exists for procurement savings specifically. Establishing that role, with a defined remit and a defined turnaround, is a precondition for everything else in this article. Where the request goes to finance generally rather than to a named individual, it will be handled by whoever has capacity, the answer will be inconsistent between categories, and the process will lapse within two quarters.

There is a variant of the signature test worth using where a full validation process is not yet in place, which is to ask the finance counterpart a narrower question: if this saving is real, which account will show it, and by when. That question is answerable in a few minutes, it does not require a commitment, and it surfaces the same information. Where no account can be named, the saving is not a hard saving whatever else it may be, and the conversation moves immediately to what category it belongs in.

The narrower question also has the advantage of being diagnostic rather than adversarial. A controller asked to sign is being asked to accept an obligation, and may decline for reasons of caution. A controller asked which account will move is being asked a technical question about their own ledger, which they are generally happy to answer. Organizations building toward full validation frequently find this the easier starting point and adopt the signature once the categories have already been separated.

11 The fairness case: avoidance is real value

This article has been critical of how savings are constructed and reported, and a reader who concluded that procurement savings are largely fictional would have taken away something the evidence does not support and this article does not intend.

The strongest point is that cost avoidance is real economic value and is frequently the most value a buyer can create. In a sole-source category, or in an inflationary market where every supplier is raising prices, there is no hard saving available: the counterfactual is a larger increase, and holding it down is the entire scope of what skilled buying can achieve. A function that delivered a three percent increase in a market moving at eight has protected margin as effectively as one that negotiated a five percent reduction in a stable market, and a reporting regime that recognises only the second will drive its people toward the categories where reductions are available and away from the categories where the exposure is greatest.

The second point is that a great deal of procurement value is not financially visible for reasons that have nothing to do with measurement games. Improved supplier quality reduces failure costs that were never quantified. Better security of supply prevents a disruption whose cost is unknowable because it did not occur. Contractual protections reduce risk exposure. Supplier development improves innovation access. All of these are real, all of them are properly part of what a procurement function is for, and none of them will ever appear in a general ledger line. An organization that measures its procurement function purely on traceable spend reduction is measuring a fraction of its contribution and will get a function optimised for that fraction.

The third point follows directly and is the most important qualification to this article's argument. Demanding profit and loss traceability for everything would distort behaviour toward price-only buying, which is the failure mode that produces supply disruptions, quality problems, and adversarial supplier relationships. Practitioner sources across this field make the point consistently: what gets measured is what gets chased, and over-rewarding hard savings specifically pulls buyers toward squeezing price at the expense of the relationship qualities that determine performance. The remedy proposed here is separation of categories, not elimination of the categories that cannot be traced.

The fair synthesis is therefore narrower than the article's title suggests. The illusion is not that the value is imaginary; it is that a single number can represent value with radically different verifiability. Report hard savings as hard savings and defend them with the controller's signature. Report cost avoidance as cost avoidance, with the counterfactual stated and evidenced by an index or a documented market movement rather than by a supplier's opening position. Report soft savings as capacity created, and be explicit that converting it into financial value requires a separate management decision. Three honest numbers are worth considerably more to an organization than one contested one, and the function that reports them is in a stronger position, not a weaker one.

A fifth point in fairness concerns what happens to categories where no saving is available at all, which a savings-only reporting regime handles badly. In a category with a single qualified supplier, a regulated price, or a long-term agreement not yet due for renewal, skilled procurement work consists of managing the relationship, protecting supply, and preparing for the next negotiation. None of that produces a reportable number in the current period, and a function judged only on savings will under-resource exactly those categories.

That under-resourcing is where supply disruptions originate. The categories with concentrated supply and limited commercial leverage are simultaneously the ones offering the least savings opportunity and the greatest exposure, and a reporting regime that directs attention away from them is creating operational risk in exchange for a reported figure. Adding a qualitative risk and continuity category to the report, as section fourteen proposes, is what allows a function to justify spending time where the savings are not.

12 Reconcile rather than arbitrate

The constructive core of this article is a process rather than a definition, because the definitional argument cannot be settled in the abstract and can be settled operationally.

A reconciliation takes a reported savings figure and walks it to the ledger, adjusting explicitly for each mechanism identified in Figure 6. Start with the reported number. Remove the categories that are not hard savings and report them separately. Adjust for volume change so that the comparison is on a like-for-like basis. Adjust for mix using line-level data. Adjust for index and currency movements. Adjust for timing so that the benefit is attributed to the period in which it occurred. Identify the residual and trace it to the accounts in which it should appear. The output is a number both functions can see, plus a documented explanation of every difference.

The value of this exercise is not primarily the number it produces. It is that it converts an annual argument about credibility into a technical exercise with rules, in which each adjustment is a fact rather than an opinion. Organizations that perform it typically find that the reconciled figure is materially lower than the reported one and that the reasons are entirely mundane, at which point the conversation shifts from whether procurement is exaggerating to what the organization should do about off-contract spend, volume forecasting, or catalogue maintenance. Those are solvable operational problems and they were previously invisible behind a dispute about integrity.

The rules should be agreed once and applied consistently, which is the element organizations most often omit. A reconciliation performed with rules negotiated during the exercise will be relitigated the following year. A documented method covering baseline selection, category definitions, adjustment mechanisms, and the validation threshold can be applied by either function to any category and produces the same answer regardless of who runs it. That document is short, it takes a working group a few sessions to produce, and it is the single highest-return governance artefact available in this area.

A caution about sequencing belongs with the reconciliation proposal. Organizations that begin by reconciling a completed year, retrospectively, generally produce a large and uncomfortable gap and an argument about whose fault it is. Organizations that begin by agreeing the method prospectively, applying it to the coming year, and reconciling as they go produce a smaller gap, no argument, and a working process by the end of the period. The retrospective exercise is more satisfying to whoever proposed it and considerably less likely to survive.

The recommended sequence is therefore to agree definitions and adjustment rules first, apply them to new sourcing activity, reconcile quarterly rather than annually so that discrepancies surface while they can still be acted on, and leave prior years alone. Nothing is gained by relitigating a number that has already been reported, and the political cost of attempting it is usually enough to stop the initiative before the useful part begins.

13 A savings protocol, and a scoring rubric

The principles combine into a protocol and a rubric that a chief financial officer, an audit committee, or a chief procurement officer can apply to establish whether a savings number means anything. Figure 9 sets out the discipline.

Figure 9. Making a savings number mean one thing: agree the baseline first, report three buckets rather than one, name the finance counterpart, hold the terms constant, and ask for the signature.

The protocol runs as follows. Agree the baseline before the sourcing event begins, document it, and apply it regardless of the result. Report three separate totals for hard savings, soft savings, and cost avoidance, each with its own definition, and never publish a combined figure as the headline. Name a specific finance counterpart with authority to validate and a defined turnaround. Hold the terms constant by adjusting explicitly for volume, mix, index, currency, and timing according to a documented method agreed in advance. And apply the signature test to the hard-savings number, treating a refusal as information about the saving rather than as an obstacle to be escalated around.

A scoring rubric

The dimensions below distinguish a validated savings number from a reported one.

Dimension	Validated	Reported only
Baseline	Agreed and documented before sourcing	Selected when the result is written up
Categories	Three separate totals, defined	One combined headline figure
Adjustments	Volume, mix, index, currency, timing	None, or applied selectively
Counterfactual basis	Index or documented market movement	The supplier's opening position
Finance role	Named counterpart validates before publication	Informed after the number is set
Budget treatment	Hard savings reflected in the reforecast	No corresponding budget movement
Benchmark use	Own reconciled trend over time	Interested industry realization figures

A number scoring in the left column can be relied upon in planning and defended to an audit committee. One scoring in the right column will be discounted by finance at an arbitrary rate, which penalises the traceable portion along with the rest. The rubric does not reduce what procurement claims to have achieved; it separates the claim into parts that can each be argued on their own evidence.

14 What to do about the categories you cannot trace

Separating the categories raises an obvious question, which is what an organization should do with the two categories that will never appear in the accounts, and the answer is not to stop reporting them.

Cost avoidance should be reported with its counterfactual stated explicitly and evidenced independently of the negotiation. Where a published index, a tariff schedule, a comparable market movement, or a documented cost-driver model supports the claim that prices in this category were rising by a given amount, the avoidance claim rests on external evidence and is strong. Where the only evidence is what the supplier initially proposed, the claim rests on a negotiating posture and should be labelled as such. Making this distinction visible within the avoidance category is more useful than arguing about whether avoidance should be counted at all.

Soft savings should be reported as capacity created rather than as money, with the unit of measurement being hours, headcount equivalent, or throughput rather than currency. Converting them to a currency figure invites a comparison with hard savings that they cannot survive and obscures what would have to happen for the value to be realised. Reported as capacity, they prompt the right question, which is whether the organization intends to redeploy that capacity or absorb it, and that is a management decision that someone should be asked to make.

Risk and quality outcomes deserve a fourth category that most organizations do not maintain at all, reported qualitatively with supporting measures rather than in currency. Supplier failure rates, quality incidents, delivery reliability, contractual protections obtained, and reductions in single-source exposure are all measurable and none converts sensibly into a savings figure. Reporting them alongside the financial categories gives the function credit for work that a savings-only regime ignores entirely, and it counteracts the distortion toward price-only buying that section eleven identified as the main risk of this article's own recommendations.

The general principle is that each kind of value should be reported in the unit that suits it and defended with the evidence appropriate to it. Currency for what is traceable in currency, capacity for what is capacity, external evidence for counterfactual claims, and operational measures for risk and quality. This produces a longer report than a single headline figure and it produces one that survives scrutiny. Functions that have made this change generally report that the annual review becomes a discussion of what to do next rather than an argument about whether the number is real, which is the outcome worth having.

A final point concerns how these four categories should be weighted in whatever objective the function is set, because reporting them separately achieves little if the target remains a single blended number. Where the target is expressed as one figure, the function will optimise the categories that are easiest to grow, which are the ones with the weakest verification. Where each category carries its own objective, the incentive to migrate value between them disappears.

Setting separate objectives also permits a more sophisticated conversation about what the organization actually wants in a given year. A business under margin pressure may reasonably weight hard savings heavily. A business facing supply concentration may weight risk and continuity outcomes. Making that emphasis explicit through the objective structure is a legitimate management choice, and it is unavailable to an organization whose only instrument is a single savings percentage.

15 Conclusion: credibility is the asset

Procurement functions create substantial value and report it in a form that invites disbelief. That is the argument of this article, and the three mechanisms behind it are specific: baselines chosen by the party reporting the result, cost avoidance measured against counterfactuals that are sometimes evidenced and sometimes are the supplier's opening position, and the absence of any validation by a function with a different incentive. None of these requires bad faith and all of them produce an upward bias that finance detects and responds to by discounting the whole figure.

The evidence position deserves restating because it is unusual. This research found no independent, sampled, peer-reviewed study of what proportion of reported procurement savings reaches the financial statements. Every realization figure in general circulation originates with a consultancy selling procurement advisory work or a vendor selling savings-tracking software, and the methods behind those figures are not disclosed. The only disinterested evidence comes from public-sector audit, which has documented cost reduction and cost avoidance being reported as one blended figure, and which examines specific programmes rather than sampling a population. An organization therefore cannot benchmark its realization rate against anything reliable, and should measure its own.

What this does not support is the conclusion that procurement savings are fictional. Cost avoidance in a sole-source or inflationary category is frequently the most value a buyer can create and is invisible by construction. Quality, security of supply, and risk reduction are real contributions that no ledger records. Demanding financial traceability for all of it would drive the function toward price-only buying, which is the behaviour that produces the disruptions and quality failures that cost far more than the savings gained. The problem is the single combined number, not the existence of value that a general ledger cannot see.

What a chief procurement officer or a chief financial officer should do is available now and requires no system, no vendor, and no budget. Agree the baseline before the sourcing event rather than when the result is written up. Report hard savings, cost avoidance, and capacity created as three separate totals with three separate definitions, and add a fourth qualitative category for risk and quality. Name a finance counterpart with authority to validate and a defined turnaround, and build a documented adjustment method covering volume, mix, index, currency, and timing that either function can apply to any category and get the same answer. Then apply the signature test to the hard-savings number and treat a refusal as information rather than as an obstacle. The number that emerges will be smaller than the one currently reported and it will be believed, and a smaller number that finance will plan against is worth considerably more to an organization than a larger one that it discounts. Credibility, not magnitude, is the asset that a procurement function should be managing.

16 Methodology, caveats, and sources

Methodology

- This article draws on public-sector audit reports, practitioner and vendor literature, and peer-reviewed venue review, current to mid-2026. Supply Chain Research is independent and accepts no payment from the consultancies, source-to-pay vendors, or spend-analytics providers discussed.
- Every realization-rate and benchmark figure cited is attributed to its publisher with the commercial interest stated. Where no disinterested figure exists, that absence is reported as a finding rather than resolved by selecting an interested estimate.

Caveats

- No independent, sampled, peer-reviewed study of procurement savings realization was located by this research. All quantified realization figures in this article originate with parties selling procurement advisory services or savings-tracking software and are identified as such wherever used.
- The stage-cascade percentages are published by a spend-analytics vendor without disclosed method or sample. They are presented as the field's description of itself and should not be treated as measurement.
- Benchmark figures concerning world-class procurement performance derive from a proprietary benchmarking panel maintained by a consultancy that monetises both the benchmark and the associated advisory work.
- This research did not complete a systematic review of non-United States national audit office value-for-money reporting on claimed procurement and efficiency savings. Readers seeking additional independent evidence should consult those offices directly. This gap is stated rather than filled.
- Figures 1, 3, 5, 6, 7, and 9 are conceptual or illustrative presentations of structure rather than measured data. Figures 4 and 8 are explicitly qualitative assessments and present no measured data.
- This article addresses the construction, categorisation, and validation of reported savings. It is not accounting, audit, tax, or legal advice, and it does not assess any particular software product or advisory firm.

Sources

[1] US Government Accountability Office. Contract audits: role in helping ensure effective oversight and reducing improper payments, GAO-11-331T.

[2] US Government Accountability Office. Defense acquisitions: Navy and Marine Corps pilot program initiatives to reduce total ownership costs, GAO-01-675R.

[3] US Government Accountability Office. 2022 annual report on fragmentation, overlap, and duplication, GAO-22-105301.

[4] Journal of Purchasing and Supply Management (Elsevier). Peer-reviewed venue for purchasing and supply management research, including savings measurement and value leakage.

[5] Suplari. Cost savings versus cost avoidance, including the four-stage leakage model (interested source: spend-analytics vendor).

[6] Simfoni. Glossary entry on cost avoidance and its evidentiary limitations (interested source: source-to-pay vendor).

[7] Stampli. What counts as procurement savings and how to prove they reach the profit and loss account (interested source: vendor).

[8] ProcurementAIAgents. Comparison of cost savings and cost avoidance, including the behavioural effects of metric selection (interested source: vendor).

Additional context drawn from Hackett Group benchmark publications, which are identified as interested sources throughout, and from practitioner accounts of savings validation practice. Readers seeking independent non-United States evidence are directed to national audit office value-for-money reporting, which this research did not systematically review. This article is analysis, not accounting, audit, tax, or legal advice, and its conclusions should be validated against your own circumstances and reporting obligations before any decision.